P1 1127612

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

February 25, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: ***60/431,133***
FILING DATE: ***December 04, 2002***
RELATED PCT APPLICATION NUMBER: *PCT/US03/38808*

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

P. SWAIN
Certifying Officer

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

12-06-02

PTO/SB/16 (10-01)
Approved for use through 10/31/2002. OMB 0651-0032
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53 (c).

Express Mail Label No. ET 657 238 023 US

INVENTOR(S)		
Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Roberto	Macina	San Jose, CA
Leah R.	Turner	Sunnyvale, CA
Yongming	Sun	Redwood City, CA

☐ *Additional inventors are being named on the _____ separately numbered sheets attached hereto*

TITLE OF THE INVENTION (280 characters max)

COMPOSITIONS AND METHODS RELATING TO COLON SPECFIFIC GENES AND PROTEINS

CORRESPONDENCE ADDRESS

Direct all correspondence to:

☐ Customer Number ☐ → *Place Customer Number Bar Code Label here*

OR *Type Customer Number here*

☒ Firm or Individual Name	Licata & Tyrrell				
Address	66 East Main Street				
Address					
City	Marlton	State	New Jersey	ZIP	08053
Country	USA	Telephone	(805) 810-1515	Fax	(805) 810-1454

ENCLOSED APPLICATION PARTS *(check all that apply)*

☒ Specification *Total Number of Pages* 262 ☐ CD(s), Number ☐

☐ Drawing(s) *Number of Sheets* ☐ ☐ Other (specify) ☐

☐ Application Data Sheet. See 37 CFR 1.76

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

☐ Applicant claims small entity status. See 37 CFR 1.27.

☐ A check or money order is enclosed to cover the filing fees — FILING FEE AMOUNT ($)

☒ The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 50-0434 — 160.00

☐ Payment by credit card. Form PTO-2038 is attached.

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number are: _____.

Respectfully submitted,

SIGNATURE _____ Date: 4 December 2002

TYPED or PRINTED NAME Nathan P. Letts REGISTRATION NO. *(if appropriate)* 36,581

TELEPHONE (650) 246-6400 Docket Number: DEX-0356

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

This collection of information is required by 37 CFR 1.51. The information is used by the public to file (and by the PTO to process) a provisional application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the complete provisional application to the PTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Washington, D.C., 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231

DEX – 0356

CERTIFICATE OF EXPRESS MAILING

"Express Mail" Label No. ET 657 238 023 US

Date of Deposit: 4 December 2002

I hereby certify that this paper is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R 1.10 on the date indicated above and is addressed to the "BOX PROVISIONAL", Assistant Commissioner for Patents, Washington, D.C. 20231.

1) Provision Application Transmittal Letter, with authorization to charge deposit account 50-0434 $160.00 for filing fee (2 copies);

2) Application consisting of 262 pages including; 256 pages of Specification, 5 pages of Claims, and 1 page of Abstract; and

3) Return Post Card.

NATHAN P. LETTS
Registration No. 36,581

COMPOSITIONS AND METHODS RELATING TO COLON SPECFIFIC GENES AND PROTEINS

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic colon cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, post translational modifications (PTMs), variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and non-cancerous disease states in colon, identifying colon tissue and monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, therapeutic molecules including but limited to antibodies or antisense molecules, production of transgenic animals and cells, and production of engineered colon tissue for treatment and research.

BACKGROUND OF THE INVENTION

Colorectal cancer is the second most common cause of cancer death in the United States and the third most prevalent cancer in both men and women. M. L. Davila & A. D. Davila, *Screening for Colon and Rectal Cancer, in* Colon and Rectal Cancer 47 (Peter S. Edelstein ed., 2000). The American Cancer Society estimates that there will be about 107,300 new cases of colon cancer and 41,000 new cases of rectal cancer in 2002 in the United States. Colon cancer will cause about 48,100 deaths and rectal cancer about 8,500 deaths. ACS Website: http://www.cancer.org. Nearly all cases of colorectal cancer arise from adenomatous polyps, some of which mature into large polyps, undergo abnormal growth and development, and ultimately progress into cancer. Davila at 55-56. This progression would appear to take at least 10 years in most patients, rendering it a readily treatable form of cancer if diagnosed early, when the cancer is localized. Davila at 56; Walter J. Burdette, Cancer: Etiology, Diagnosis, and Treatment 125 (1998).

Although our understanding of the etiology of colon cancer is undergoing continual refinement, extensive research in this area points to a combination of factors,

including age, hereditary and nonheriditary conditions, and environmental/dietary factors. Age is a key risk factor in the development of colorectal cancer, Davila at 48, with men and women over 40 years of age become increasingly susceptible to that cancer, Burdette at 126. Incidence rates increase considerably in each subsequent decade of life. Davila at 48. A number of hereditary and nonhereditary conditions have also been linked to a heightened risk of developing colorectal cancer, including familial adenomatous polyposis (FAP), hereditary nonpolyposis colorectal cancer (Lynch syndrome or HNPCC), a personal and/or family history of colorectal cancer or adenomatous polyps, inflammatory bowel disease, diabetes mellitus, and obesity. *Id.* at 47; Henry T. Lynch & Jane F. Lynch, *Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndromes), in* Colon and Rectal Cancer 67-68 (Peter S. Edelstein ed., 2000).

Environmental/dietary factors associated with an increased risk of colorectal cancer include a high fat diet, intake of high dietary red meat, and sedentary lifestyle. Davila at 47; Reddy, B. S., *Prev. Med.* 16(4): 460-7 (1987). Conversely, environmental/dietary factors associated with a reduced risk of colorectal cancer include a diet high in fiber, folic acid, calcium, and hormone-replacement therapy in post-menopausal women. Davila at 50-55. The effect of antioxidants in reducing the risk of colon cancer is unclear. Davila at 53.

Because colon cancer is highly treatable when detected at an early, localized stage, screening should be a part of routine care for all adults starting at age 50, especially those with first-degree relatives with colorectal cancer. One major advantage of colorectal cancer screening over its counterparts in other types of cancer is its ability to not only detect precancerous lesions, but to remove them as well. Davila at 56. The key colorectal cancer screening tests in use today are fecal occult blood test, sigmoidoscopy, colonoscopy, double-contrast barium enema, and the carcinoembryonic antigen (CEA) test. Burdette at 125; Davila at 56.

The fecal occult blood test (FOBT) screens for colorectal cancer by detecting the amount of blood in the stool, the premise being that neoplastic tissue, particularly malignant tissue, bleeds more than typical mucosa, with the amount of bleeding increasing with polyp size and cancer stage. Davila at 56-57. While effective at detecting early stage tumors, FOBT is unable to detect adenomatous polyps (premalignant lesions), and, depending on the contents of the fecal sample, is subject to rendering false positives. Davila at 56-59. Sigmoidoscopy and colonoscopy, by contrast, allow direct visualization

of the bowel, and enable one to detect, biopsy, and remove adenomatous polyps. Davila at 59-60, 61. Despite the advantages of these procedures, there are accompanying downsides: sigmoidoscopy, by definition, is limited to the sigmoid colon and below, colonoscopy is a relatively expensive procedure, and both share the risk of possible bowel perforation and hemorrhaging. Davila at 59-60. Double-contrast barium enema (DCBE) enables detection of lesions better than FOBT, and almost as well a colonoscopy, but it may be limited in evaluating the winding rectosigmoid region. Davila at 60. The CEA blood test, which involves screening the blood for carcinoembryonic antigen, shares the downside of FOBT, in that it is of limited utility in detecting colorectal cancer at an early stage. Burdette at 125.

Once colon cancer has been diagnosed, treatment decisions are typically made in reference to the stage of cancer progression. A number of techniques are employed to stage the cancer (some of which are also used to screen for colon cancer), including pathologic examination of resected colon, sigmoidoscopy, colonoscopy, and various imaging techniques. AJCC Cancer Staging Handbook 84 (Irvin D. Fleming et al. eds., 5th ed. 1998); Montgomery, R. C. and Ridge, J.A., *Semin. Surg. Oncol.* 15(3): 143-150 (1998). Moreover, chest films, liver functionality tests, and liver scans are employed to determine the extent of metastasis. Fleming at 84. While computerized tomography and magnetic resonance imaging are useful in staging colorectal cancer in its later stages, both have unacceptably low staging accuracy for identifying early stages of the disease, due to the difficulty that both methods have in (1) revealing the depth of bowel wall tumor infiltration and (2) diagnosing malignant adenopathy. Thoeni, R. F., *Radiol. Clin. N. Am.* 35(2): 457-85 (1997). Rather, techniques such as transrectal ultrasound (TRUS) are preferred in this context, although this technique is inaccurate with respect to detecting small lymph nodes that may contain metastases. David Blumberg & Frank G. Opelka, *Neoadjuvant and Adjuvant Therapy for Adenocarcinoma of the Rectum, in* Colon and Rectal Cancer 316 (Peter S. Edelstein ed., 2000).

Several classification systems have been devised to stage the extent of colorectal cancer, including the Dukes' system and the more detailed International Union against Cancer-American Joint Committee on Cancer TNM staging system, which is considered by many in the field to be a more useful staging system. Burdette at 126-27. The TNM system, which is used for either clinical or pathological staging, is divided into four stages, each of which evaluates the extent of cancer growth with respect to primary tumor (T),

regional lymph nodes (N), and distant metastasis (M). Fleming at 84-85. The system focuses on the extent of tumor invasion into the intestinal wall, invasion of adjacent structures, the number of regional lymph nodes that have been affected, and whether distant metastasis has occurred. Fleming at 81.

Stage 0 is characterized by *in situ* carcinoma (Tis), in which the cancer cells are located inside the glandular basement membrane (intraepithelial) or lamina propria (intramucosal). In this stage, the cancer has not spread to the regional lymph nodes (N0), and there is no distant metastasis (M0). In stage I, there is still no spread of the cancer to the regional lymph nodes and no distant metastasis, but the tumor has invaded the submucosa (T1) or has progressed further to invade the muscularis propria (T2). Stage II also involves no spread of the cancer to the regional lymph nodes and no distant metastasis, but the tumor has invaded the subserosa, or the nonperitonealized pericolic or perirectal tissues (T3), or has progressed to invade other organs or structures, and/or has perforated the visceral peritoneum (T4). Stage III is characterized by any of the T substages, no distant metastasis, and either metastasis in 1 to 3 regional lymph nodes (N1) or metastasis in four or more regional lymph nodes (N2). Lastly, stage IV involves any of the T or N substages, as well as distant metastasis. Fleming at 84-85; Burdette at 127.

Currently, pathological staging of colon cancer is preferable over clinical staging as pathological staging provides a more accurate prognosis. Pathological staging typically involves examination of the resected colon section, along with surgical examination of the abdominal cavity. Fleming at 84. Clinical staging would be a preferred method of staging were it at least as accurate as pathological staging, as it does not depend on the invasive procedures of its counterpart.

Turning to the treatment of colorectal cancer, surgical resection results in a cure for roughly 50% of patients. Irradiation is used both preoperatively and postoperatively in treating colorectal cancer. Chemotherapeutic agents, particularly 5-fluorouracil, are also powerful weapons in treating colorectal cancer. Other agents include irinotecan and floxuridine, cisplatin, levamisole, methotrexate, interferon-α, and leucovorin. Burdette at 125, 132-33. Nonetheless, thirty to forty percent of patients will develop a recurrence of colon cancer following surgical resection, which in many patients is the ultimate cause of death. Wayne De Vos, *Follow-up After Treatment of Colon Cancer*, Colon and Rectal Cancer 225 (Peter S. Edelstein ed., 2000). Accordingly, colon cancer patients must be

closely monitored to determine response to therapy and to detect persistent or recurrent disease and metastasis.

The next few paragraphs describe the some of molecular bases of colon cancer. In the case of FAP, the tumor suppressor gene APC (adenomatous polyposis coli), chromosomally located at 5q21, has been either inactivated or deleted by mutation. Alberts et al., Molecular Biology of the Cell 1288 (3d ed. 1994). The APC protein plays a role in a number of functions, including cell adhesion, apoptosis, and repression of the *c-myc* oncogene. N. R. Hall & R. D. Madoff, *Genetics and the Polyp-Cancer Sequence*, Colon and Rectal Cancer 8 (Peter S. Edelstein, ed., 2000). Of those patients with colorectal cancer who have normal APC genes, over 65% have such mutations in the cancer cells but not in other tissues. Alberts et al., *supra* at 1288. In the case of HPNCC, patients manifest abnormalities in the tumor suppressor gene HNPCC, but only about 15% of tumors contain the mutated gene. *Id.* A host of other genes have also been implicated in colorectal cancer, including the K-*ras*, N-*ras*, H-*ras* and c-*myc* oncogenes, and the tumor suppressor genes *DCC* (deleted in colon carcinoma) and *p53*. Hall & Madoff, *supra* at 8-9; Alberts et al., *supra* at 1288.

Abnormalities in Wg/Wnt signal transduction pathway are also associated with the development of colorectal carcinoma. Taipale, J. and Beachy, P.A. *Nature* 411: 349-354 (2001). Wnt1 is a secreted protein gene originally identified within mouse mammary cancers by its insertion into the mouse mammary tumor virus (MMTV) gene. The protein is homologous to the wingless (Wg) gene product of Drosophila, in which it functions as an important factor for the determination of dorsal-ventral segmentation and regulates the formation of fly imaginal discs. Wg/Wnt pathway controls cell proliferation, death and differentiation. Taipal (2001). There are at least 13 members in the Wnt family. These proteins have been found expressed mainly in the central nervous system (CNS) of vertebrates as well as other tissues such as mammary and intestine. The Wnt proteins are the ligands for a family of seven transmembrane domain receptors related to the Frizzled gene product in Drosophila. Binding Wnt to Frizzled stimulates the activity of the downstream target, Dishevelled, which in turn inactivates the glycogen synthesase kinase 3β (GSK3β). Taipal (2001). Usually active GSK3β will form a complex with the adenomatous polyposis coli (APC) protein and phosphorylate another complex member, β-catenin. Once phosphorylated, β-catenin is directed to degradation through the ubiquitin pathway. When GSK3β or APC activity is down regulated, β-catenin is accumulated in

the cytoplasm and binds to the T-cell factor or lymphocyte excitation factor (Tcf/Lef) family of transcriptional factors. Binding of β-catenin to Tcf releases the transcriptional repression and induces gene transcription. Among the genes regulated by β-catenin are a transcriptional repressor Engrailed, a transforming growth factor-β (TGF-β) family member Decapentaplegic, and the cytokine Hedgehog in Drosophila. β-Catenin also involves in regulating cell adhesion by binding to α-catenin and E-cadherin. On the other hand, binding of β-catenin to these proteins controls the cytoplasmic β-catenin level and its complexing with TCF. Taipal (2001). Growth factor stimulation and activation of c-src or v-src also regulate β-catenin level by phosphorylation of α-catenin and its related protein, $p120^{cas}$. When phosphorylated, these proteins decrease their binding to E-cadherin and β-catenin resulting in the accumulation of cytoplasmic β-catenin. Reynolds, A.B. et al. *Mol. Cell Biol.* 14: 8333-8342 (1994). In colon cancer, c-src enzymatic activity has been shown increased to the level of v-src. Alternation of components in the Wg/Wnt pathway promotes colorectal carcinoma development. The best known modifications are to the APC gene. Nicola S et al. *Hum. Mol. Genet* 10:721-733 (2001). This germline mutation causes the appearance of hundreds to thousands of adenomatous polyps in the large bowel. It is the gene defect that accounts for the autosomally dominantly inherited FAP and related syndromes. The molecular alternations that occur in this pathway largely involve deletions of alleles of tumor-suppressor genes, such as APC, p53 and Deleted in Colorectal Cancer (DCC), combined with mutational activation of proto-oncogenes, especially c-Ki-ras. Aoki, T. et al. *Human Mutat.* 3: 342-346 (1994). All of these lead to genomic instability in colorectal cancers.

Another source of genomic instability in colorectal cancer is the defect of DNA mismatch repair (MMR) genes. Human homologues of the bacterial *mut*HLS complex (hMSH2, hMLH1, hPMS1, hPMS2 and hMSH6), which is involved in the DNA mismatch repair in bacteria, have been shown to cause the HNPCC (about 70-90% HNPCC) when mutated. Modrich, P. and Lahue, R. *Ann Rev. Biochem.* 65: 101-133 (1996); and Peltomäki, P. *Hum. Mol. Genet* 10: 735-740 (2001). The inactivation of these proteins leads to the accumulation of mutations and causes genetic instability that represents errors in the accurate replication of the repetitive mono-, di-, tri- and tetra-nucleotide repeats, which are scattered throughout the genome (microsatellite regions). Jass, J.R. et al. *J. Gastroenterol Hepatol* 17: 17-26 (2002). Like in the classic FAP, mutational activation of

c-Ki-ras is also required for the promotion of MSI in the alternative HNPCC. Mutations in other proteins such as the tumor suppressor protein phosphatase PTEN (Zhou, X.P. et al. *Hum. Mol. Genet* 11: 445-450 (2002)), BAX (Buttler, L.M. *Aus. N. Z. J. Surg.* 69: 88-94 (1999)), Caspase-5 (Planck, M. *Cancer Genet Cytogenet.* 134: 46-54 (2002)), TGFβ-RII (Fallik, D. et al. *Gastroenterol Clin Biol.* 24: 917-22 (2000)) and IGFII-R (Giovannucci E. *J. Nutr.* 131: 3109S-20S (2001)) have also been found in some colorectal tumors possibly as the cause of MMR defect.

Some tyrosine kinases have been shown up-regulated in colorectal tumor tissues or cell lines like HT29. Skoudy, A. et al. *Biochem J.* 317 (Pt 1): 279-84 (1996). Focal adhesion kinase (FAK) and its up-stream kinase c-src and c-yes in colonic epithelia cells may play an important role in the promotion of colorectal cancers through the extracellular matrix (ECM) and integrin-mediated signaling pathways. Jessup, J.M. et al., *The molecular biology of colorectal carcinoma, in:* The Molecular Basis of Human Cancer, 251-268 (Coleman W.B. and Tsongalis G.J. Eds. 2002). The formation of c-src/FAK complexes may coordinately deregulate VEGF expression and apoptosis inhibition. Recent evidences suggest that a specific signal-transduction pathway for cell survival that implicates integrin engagement leads to FAK activation and thus activates PI-3 kinase and akt. In turn, akt phosphorylates BAD and blocks apoptosis in epithelial cells. The activation of c-src in colon cancer may induce VEGF expression through the hypoxia pathway. Other genes that may be implicated in colorectal cancer include Cox enzymes (Ota, S. et al. *Aliment Pharmacol. Ther.* 16 (Suppl 2): 102-106 (2002)), estrogen (al-Azzawi, F. and Wahab, M. *Climacteric* 5: 3-14 (2002)), peroxisome proliferator-activated receptor-γ (PPAR-γ) (Gelman, L. et al. *Cell Mol. Life Sci.* 55: 932-943 (1999)), IGF-I (Giovannucci (2001)), thymine DNA glycosylase (TDG) (Hardeland, U. et al. *Prog. Nucleic Acid Res. Mol. Biol.* 68: 235-253 (2001)) and EGF (Mendelsohn, J. *Endocrine-Related Cancer* 8: 3-9 (2001)).

Gene deletion and mutation are not the only causes for development of colorectal cancers. Epigenetic silencing by DNA methylation also accounts for the lost of function of colorectal cancer suppressor genes. A strong association between MSI and CpG island methylation has been well characterized in sporadic colorectal cancers with high MSI but not in those of hereditary origin. In one experiment, DNA methylation of MLH1, CDKN2A, MGMT, THBS1, RARB, APC, and p14ARF genes has been shown in 80%, 55%, 23%, 23%, 58%, 35%, and 50% of 40 sporadic colorectal cancers with high MSI

respectively. Yamamoto, H. et al. *Genes Chromosomes Cancer* 33: 322-325 (2002); and Kim, K.M. et al. *Oncogene.* 12;21(35): 5441-9 (2002). Carcinogen metabolism enzymes such as GST, NAT, CYP and MTHFR are also associated with an increased or decreased colorectal cancer risk. Pistorius, S. et al. *Kongressbd Dtsch Ges Chir Kongr* 118: 820-824 (2001); and Potter, J.D. *J. Natl. Cancer Inst.* 91: 916-932 (1999).

From the foregoing, it is clear that procedures used for detecting, diagnosing, monitoring, staging, prognosticating, and preventing the recurrence of colorectal cancer are of critical importance to the outcome of the patient. Moreover, current procedures, while helpful in each of these analyses, are limited by their specificity, sensitivity, invasiveness, and/or their cost. As such, highly specific and sensitive procedures that would operate by way of detecting novel markers in cells, tissues, or bodily fluids, with minimal invasiveness and at a reasonable cost, would be highly desirable.

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop colorectal cancer, for diagnosing colorectal cancer, for monitoring the progression of the disease, for staging the colorectal cancer, for determining whether the colorectal cancer has metastasized, and for imaging the colorectal cancer. Following accurate diagnosis, there is also a need for less invasive and more effective treatment of colorectal cancer.

SUMMARY OF THE INVENTION

The present invention solves many needs in the art by providing nucleic acid molecules, polypeptides and antibodies thereto, variants and derivatives of the nucleic acids and polypeptides, agonists and antagonists that may be used to identify, diagnose, monitor, stage, image and treat colon cancer and non-cancerous disease states in colon; identify and monitor colon tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered colon tissue for treatment and research.

One aspect of the present invention relates to nucleic acid molecules that are specific to colon cells, colon tissue and/or the colon organ. These colon specific nucleic acids (CSNAs) may be a naturally occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally occurring nucleic acid molecule. If the CSNA is genomic DNA, then the CSNA is a colon specific gene (CSG). If the

CSNA is RNA, then it is a colon specific transcript encoded by a CSG. Due to alternative splicing and transcriptional modification one CSG may encode for multiple colon specific RNAs. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to colon. More preferred is a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 101-194. In another preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1-100. For the CSNA sequences listed herein, DEX0356_001.nt.1 corresponds to SEQ ID NO: 1. For sequences with multiple splice variants, the parent sequence DEX0356_001.nt.1, will be followed by DEX0356_001.nt.2, etc. for each splice variant. The sequences off the corresponding peptides are listed as DEX0356_001.aa.1, etc. For the mapping of all of the nucleotides and peptides, see the table in the Example 1 section below.

This aspect of the present invention also relates to nucleic acid molecules that selectively hybridize or exhibit substantial sequence similarity to nucleic acid molecules encoding a colon Specific Protein (CSP), or that selectively hybridize or exhibit substantial sequence similarity to a CSNA. In one embodiment of the present invention the nucleic acid molecule comprises an allelic variant of a nucleic acid molecule encoding a CSP, or an allelic variant of a CSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid sequence that encodes a CSP or a part of a nucleic acid sequence of a CSNA.

In addition, this aspect of the present invention relates to a nucleic acid molecule further comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of a CSNA or the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of a CSP.

Another aspect of the present invention relates to vectors and/or host cells comprising a nucleic acid molecule of this invention. In a preferred embodiment, the nucleic acid molecule of the vector and/or host cell encodes all or a fragment of a CSP. In another preferred embodiment, the nucleic acid molecule of the vector and/or host cell comprises all or a part of a CSNA. Vectors and host cells of the present invention are useful in the recombinant production of polypeptides, particularly CSPs of the present invention.

Another aspect of the present invention relates to polypeptides encoded by a nucleic acid molecule of this invention. The polypeptide may comprise either a fragment

or a full-length protein. In a preferred embodiment, the polypeptide is a CSP. However, this aspect of the present invention also relates to mutant proteins (muteins) of CSPs, fusion proteins of which a portion is a CSP, and proteins and polypeptides encoded by allelic variants of a CSNA as provided herein.

Another aspect of the present invention relates to antibodies and other binders that specifically binds to a polypeptide of the instant invention. Accordingly antibodies or binders of the present specifically bind to CSPs, muteins, fusion proteins, and/or homologous proteins or a polypeptides encoded by allelic variants of an CSNA as provided herein.

Another aspect of the present invention relates to agonists and antagonists of the nucleic acid molecules and polypeptides of this invention. The agonists and antagonists of the instant invention may be used to treat colon cancer and non-cancerous disease states in colon and to produce engineered colon tissue.

Another aspect of the present invention relates to methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. Such methods are useful in identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and non-cancerous disease states in colon. Such methods are also useful in identifying and/or monitoring colon tissue. In addition, measurement of levels of one or more of the nucleic acid molecules of this invention may be useful for diagnostics as part of panel in combination with known other markers, particularly those described in the colon cancer background section above.

Another aspect of the present invention relates to use of the nucleic acid molecules of this invention in gene therapy, for producing transgenic animals and cells, and for producing engineered colon tissue for treatment and research.

Another aspect of the present invention relates to methods for detecting polypeptides this invention, preferably using antibodies thereto. Such methods are useful to identify, diagnose, monitor, stage, image and treat colon cancer and non-cancerous disease states in colon. In addition, measurement of levels of one or more of the polypeptides of this invention may be useful to identify, diagnose, monitor, stage, image colon cancer in combination with known other markers, particularly those described in the colon cancer background section above. The polypeptides of the present invention can

also be used to identify and/or monitor colon tissue, and to produce engineered colon tissue.

Yet another aspect of the present invention relates to a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences. In addition, the computer records regarding the nucleic acid and/or amino acid sequences and/or measurements of their levels may be used alone or in combination with other markers to diagnose colon related diseases.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999).

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described

herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single and double stranded forms of DNA. In addition, a polynucleotide may include either or both naturally occurring and modified nucleotides linked together by naturally occurring and/or non-naturally occurring nucleotide linkages.

Nucleotides are represented by single letter symbols in nucleic acid molecule sequences. The following table lists symbols identifying nucleotides or groups of nucleotides which may occupy the symbol position on a nucleic acid molecule. *See* Nomenclature Committee of the International Union of Biochemistry (NC-IUB), Nomenclature for incompletely specified bases in nucleic acid sequences, Recommendations 1984., *Eur J Biochem.* 150(1):1-5 (1985).

Symbol	Meaning	Group/Origin of Designation	Complementary Symbol
a	a	Adenine	t/u
g	g	Guanine	c
c	c	Cytosine	g
t	t	Thymine	a
u	u	Uracil	a
r	g or a	puRine	y
y	t/u or c	pYrimidine	r
m	a or c	aMino	k
k	g or t/u	Keto	m
s	g or c	Strong interactions 3H-bonds	w
w	a or t/u	Weak interactions 2H-bonds	s
b	g or c or t/u	not a	v
d	a or g or t/u	not c	h
h	a or c or t/u	not g	d
v	a or g or c	not t, not u	b

n	a or g or c or t/u, unknown, or other	aNy	n

The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the

nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and U.S. Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic

acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single-or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including in vitro recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally occurring nucleotide" referred to herein includes naturally occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such

as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phoshoraniladate, phosphoroamidate, and the like. *See e.g.*, LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), and U.S. Patent No. 5,151,510, the disclosure of which is hereby incorporated by reference in its entirety.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98

(1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998)). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, double stranded RNA (dsRNA) inhibition (RNAi), combination of triplex and antisense, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology" interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists between a first and second nucleic acid sequence when the first nucleic acid sequence or fragment thereof hybridizes to an antisense strand of the second nucleic acid, under selective hybridization conditions. Typically, selective hybridization will occur between the first nucleic acid sequence and an antisense strand of the second nucleic acid sequence when there is at least about 55% sequence identity between the first and second nucleic acid sequences— preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% — over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. *See* Sambrook (1989), *supra*, p. 9.51.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$T_m = 81.5°C + 16.6 (\log_{10}[Na^+]) + 0.41$ (fraction G + C) –

0.63 (% formamide) - (600/l) where l is the length of the hybrid in base pairs.

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5 (\log_{10}[Na^+]) + 0.58$ (fraction G + C) +

11.8 (fraction G + C)^2 - 0.35 (% formamide) - (820/l).

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5(\log_{10}[Na^+]) + 0.58$ (fraction G + C) +

11.8 (fraction G + C)^2 - 0.50 (% formamide) - (820/l).

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other

higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well known in the art.

An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (*e.g.* 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acids that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid is

created synthetically or recombinantly using a high codon degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula:

$T_m = 81.5°C + 16.6(\log_{10}[Na^+]) + 0.41(\text{fraction G+C}) - (600/N)$, wherein N is change length and the $[Na^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well known in the art. *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 μg of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 μl of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs. Techniques for ligation are well known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to

transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies. In another aspect, the invention is directed to single exon probes based on the CSNAs disclosed herein.

In one embodiment, the term "microarray" refers to a "nucleic acid microarray" having a substrate-bound plurality of nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). Additionally, these nucleic acid microarrays include substrate-bound plurality of nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000). Examples of nucleic acid microarrays may be found in U.S. Patent Nos. 6,391,623, 6,383,754, 6,383,749, 6,380,377, 6,379,897, 6,376,191, 6,372,431, 6,351,712 6,344,316, 6,316,193, 6,312,906, 6,309,828, 6,309,824, 6,306,643, 6,300,063, 6,287,850, 6,284,497, 6,284,465, 6,280,954, 6,262,216, 6,251,601, 6,245,518, 6,263,287, 6,251,601, 6,238,866, 6,228,575, 6,214,587, 6,203,989, 6,171,797, 6,103,474, 6,083,726, 6,054,274, 6,040,138, 6,083,726, 6,004,755, 6,001,309, 5,958,342, 5,952,180, 5,936,731, 5,843,655, 5,814,454, 5,837,196, 5,436,327, 5,412,087, 5,405,783, the disclosures of which are incorporated herein by reference in their entireties.

In an alternative embodiment, a "microarray" may also refer to a "peptide microarray" or "protein microarray" having a substrate-bound collection of plurality of polypeptides, the binding to each of the plurality of bound polypeptides being separately detectable. Alternatively, the peptide microarray ,may have a plurality of binders, including but not limited to monoclonal antibodies, polyclonal antibodies, phage display binders, yeast 2 hybrid binders, aptamers, which can specifically detect the binding of the polypeptides of this invention. The array may be based on autoantibody detection to the polypeptides of this invention, see Robinson *et al., Nature Medicine* 8(3):295-301 (2002). Examples of peptide arrays may be found in WO 02/31463, WO 02/25288, WO 01/94946, WO 01/88162, WO 01/68671, WO 01/57259, WO 00/61806, WO 00/54046, WO 00/47774, WO 99/40434, WO 99/39210, WO 97/42507 and U.S. Patent Nos. 6,268,210, 5,766,960, 5,143,854, the disclosures of which are incorporated herein by reference in their entireties.

In addition, determination of the levels of the CSNA or CSP may be made in a multiplex manner using techniques described in WO 02/29109, WO 02/24959, WO 01/83502, WO01/73113, WO 01/59432, WO 01/57269, WO 99/67641, the disclosures of which are incorporated herein by reference in their entireties.

The term "mutant", "mutated", or "mutation" when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment of the present invention, the nucleic acid sequence is the wild type nucleic acid sequence encoding a CSP or is a CSNA. The nucleic acid sequence may be mutated by any method known in the art including those mutagenesis techniques described *infra.*

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al., Technique* 1: 11-15 (1989) and Caldwell *et al., PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g.*, Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g.*, Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. *See, e.g.*, Delegrave *et al.*, *Biotechnology*

Research 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993).

"Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is either contiguous with the gene of interest to control the gene of interest, or acts in *trans* or at a distance to control the gene of interest.

The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors").

In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which a recombinant expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refers to that portion of a transcript-derived nucleic acid that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence is meant to be inclusive of all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally occurring and non-naturally occurring proteins and polypeptides, as well as polypeptide fragments and polypeptide mutants, derivatives and analogs thereof. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a CSP encoded by a nucleic acid molecule of the instant invention, or a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins

from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be determined by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well known in the art for purification.

The term "fragment" when used herein with respect to polypeptides of the present invention refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length CSP. In a preferred embodiment, the fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally occurring polypeptide. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" when used herein with respect to polypeptides of the present invention refers to a polypeptide which is substantially similar in primary structural sequence to a CSP but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the CSP. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation,

formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, ^{14}C and ^{3}H, ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*.

The term "fusion protein" refers to polypeptides of the present invention coupled to a heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence that encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally occurring sequence.

Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: $--CH_2NH--$, $--CH_2S--$, $--CH_2-CH_2--$, --CH=CH--(cis and trans), $--COCH_2--$, $--CH(OH)CH_2--$, and $--CH_2SO--$, by methods well known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al., Ann. Rev. Biochem.* 61:387-418 (1992)). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

The term "mutant" or "mutein" when referring to a polypeptide of the present invention relates to an amino acid sequence containing substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a CSP. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to a CSP. In an even more preferred embodiment, a mutein exhibits

95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as GAP or BESTFIT or other variation Smith-Waterman alignment. *See*, T. F. Smith and M. S. Waterman, J. Mol. Biol. 147:195-197 (1981) and W.R. Pearson, Genomics 11:635-650 (1991).

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991).

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. *See* Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed., Sinauer Associates (1991). Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α-, α-disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ-carboxyglutamate, ε-N,N,N-trimethyllysine, ε-N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the

amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

By "homology" or "homologous" when referring to a polypeptide of the present invention it is meant polypeptides from different organisms with a similar sequence to the encoded amino acid sequence of a CSP and a similar biological activity or function. Although two polypeptides are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the polypeptides. Instead, the term "homologous" is defined to mean that the two polypeptides have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous polypeptide is one that exhibits 50% sequence similarity to CSP, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous polypeptides that exhibit 80%, 85% or 90% sequence similarity to a CSP. In a yet more preferred embodiment, a homologous polypeptide exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to polypeptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. *See, e.g.*, Pearson, *Methods Mol. Biol.* 24: 307-31 (1994).

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

1) Serine (S), Threonine (T);
2) Aspartic Acid (D), Glutamic Acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);

5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and

6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992). A "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. *See, e.g.*, GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g.*, Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997). Preferred parameters for blastp are:

Expectation value:	10 (default)
Filter:	seg (default)
Cost to open a gap:	11 (default)
Cost to extend a gap:	1 (default
Max. alignments:	100 (default)
Word size:	11 (default)
No. of descriptions:	100 (default)
Penalty Matrix:	BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Algorithms other than blastp for database searching using amino acid sequences are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', $F(ab')_2$, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. A Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; a $F(ab')_2$ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; a Fd fragment consists of the VH and CH1 domains; a Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g.*, Ward *et al.*, *Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" as used herein it is meant the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which VL and VH regions are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g.*, Bird *et al.*, *Science* 242: 423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain

and creating two antigen binding sites. *See e.g.*, Holliger *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al.*, *Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than1 μM, preferably less than100 nM and most preferably less than 10 nM.

The term "patient" includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "colon specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the colon as compared to other tissues in the body. In a preferred embodiment, a "colon specific" nucleic acid molecule or polypeptide is detected at a level that is 1.5-fold higher than any other tissue in the body. In a more preferred embodiment, the "colon specific" nucleic acid molecule or polypeptide is detected at a level that is 2-fold higher than any other tissue in the body, more preferably 5-fold higher, still more preferably at least 10-fold, 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

One aspect of the invention provides isolated nucleic acid molecules that are specific to the colon or to colon cells or tissue or that are derived from such nucleic acid molecules. These isolated colon specific nucleic acids (CSNAs) may comprise cDNA genomic DNA, RNA, or a combination thereof, a fragment of one of these nucleic acids, or may be a non-naturally occurring nucleic acid molecule. A CSNA may be derived from an animal. In a preferred embodiment, the CSNA is derived from a human or other mammal. In a more preferred embodiment, the CSNA is derived from a human or other primate. In an even more preferred embodiment, the CSNA is derived from a human.

In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to colon, a colon-specific polypeptide (CSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 101-194. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1-100. Nucleotide sequences of the instantly-described nucleic acid molecules were determined by assembling several DNA molecules from either public or proprietary databases. Some of the underlying DNA

sequences are the result, directly or indirectly, of at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Amersham Biosciences, Sunnyvale, CA, USA).

Nucleic acid molecules of the present invention may also comprise sequences that selectively hybridizes to a nucleic acid molecule encoding a CSNA or a complement or antisense thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may or may not encode a CSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a CSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 101-194. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1-100 or the antisense sequence thereof. Preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a CSP under low stringency conditions. More preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a CSP under moderate stringency conditions. Most preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a CSP under high stringency conditions. In a preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 101-194. In a more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1-100.

Nucleic acid molecules of the present invention may also comprise nucleic acid sequences that exhibit substantial sequence similarity to a nucleic acid encoding a CSP or a complement of the encoding nucleic acid molecule. In this embodiment, it is preferred that the nucleic acid molecule exhibit substantial sequence similarity to a nucleic acid molecule encoding human CSP. More preferred is a nucleic acid molecule exhibiting

substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 101-194. By substantial sequence similarity it is meant a nucleic acid molecule having at least 60% sequence identity with a nucleic acid molecule encoding a CSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 101-194, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a CSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. Most preferred in this embodiment is a nucleic acid molecule that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a CSP.

The nucleic acid molecules of the present invention are also inclusive of those exhibiting substantial sequence similarity to a CSNA or its complement. In this embodiment, it is preferred that the nucleic acid molecule exhibit substantial sequence similarity to a nucleic acid molecule having a nucleic acid sequence of SEQ ID NO: 1-100. By substantial sequence similarity it is meant a nucleic acid molecule that has at least 60% sequence identity with a CSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1-100, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. More preferred is a nucleic acid molecule that has at least 90% sequence identity with a CSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. Most preferred is a nucleic acid molecule that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a CSNA.

Nucleic acid molecules that exhibit substantial sequence similarity are inclusive of sequences that exhibit sequence identity over their entire length to a CSNA or to a nucleic acid molecule encoding a CSP, as well as sequences that are similar over only a part of its length. In this case, the part is at least 50 nucleotides of the CSNA or the nucleic acid molecule encoding a CSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant

sequence identity to that of SEQ ID NO: 101-194 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1-100. The similar nucleic acid molecule may also be a naturally occurring nucleic acid molecule from a human, when the CSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of a CSNA. In a preferred embodiment, the substantially similar nucleic acid molecule is an CSNA.

The nucleic acid molecules of the present invention are also inclusive of allelic variants of a CSNA or a nucleic acid encoding a CSP. For example, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes and the sequence determined from one individual of a species may differ from other allelic forms present within the population. More than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001) – Variants with small deletions and insertions of more than a single nucleotide are also found in the general population, and often do not alter the function of the protein. In addition, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes a CSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a CSP comprising an amino acid sequence of SEQ ID NO: 101-194. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a CSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the

nucleic acid sequence of SEQ ID NO: 1-100. Also preferred is that the allelic variant is a naturally occurring allelic variant in the species of interest, particularly human.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences comprising a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is a CSP. In a preferred embodiment, the part encodes a CSP. In one embodiment, the nucleic acid molecule comprises a part of a CSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a CSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that is an allelic variant of a CSNA. In yet another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that encodes a CSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences that encode fusion proteins, homologous proteins, polypeptide fragments, muteins and polypeptide analogs, as described *infra*.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences containing modifications of the native nucleic acid molecule. Examples of such modifications include, but are not limited to, nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that may be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

Accordingly, in one embodiment, a nucleic acid molecule may include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or

fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. The labeled nucleic acid molecules are particularly useful as hybridization probes.

Common radiolabeled analogues include those labeled with ^{33}P, ^{32}P, and ^{35}S, such as α-^{32}P-dATP, α-^{32}P-dCTP, α-^{32}P-dGTP, α-^{32}P-dTTP, α-^{32}P-3'dATP, α-^{32}P-ATP, α-^{32}P-CTP, α-^{32}P-GTP, α-^{32}P-UTP, α-^{35}S-dATP, γ-^{35}S-GTP, γ-^{33}P-dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Biosciences, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. *See* Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000).

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules of the present invention can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase

chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and Peptide Nucleic Acids (PNA) to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see* Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994). Alternatively, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g.*, Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); and U.S. Patent Nos. 5,846,726, 5,925,517, 5,925,517, 5,723,591 and 5,538,848, the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the present invention may also be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See* Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.), Oligonucleotides as Therapeutic Agents – Symposium No. 209, John Wiley & Son

Ltd (1997). Such altered internucleoside bonds are often desired for techniques or for targeted gene correction, Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000). For double stranded RNA inhibition which may utilize either natural ds RNA or ds RNA modified in its, sugar, phosphate or base, *see* Hannon, *Nature* 418(11): 244-251 (2002); Fire *et al.* in WO 99/32619; Tuschl *et al.* in US2002/0086356; Kruetzer *et al.* in WO 00/44895, the disclosures of which are incorporated herein by reference in their entirety;. For circular antisense, see Kool in U.S. Patent No. 5,426,180, the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative U.S. Patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U.S. Patent Nos. 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts. Representative U.S. patents that teach the

preparation of the above backbones include, but are not limited to, U.S. Patent Nos. 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred nucleic acid molecules, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Patent Nos. 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference in its entirety. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA). PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The Tm of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the Tm of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the Tm by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the Tm by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with

nucleobase sidechains. *See, e.g.*, Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999).

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), and U.S. Patent Nos. 5,760,012 and 5,731,181, the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acid molecules of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); and Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994). Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997); Rowley *et al.*, *Mol Med* 5(10): 693-700 (1999); Kool, *Annu Rev Biophys Biomol Struct.* 25: 1-28 (1996).

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect and characterize gross alterations in the gene of a CSNA, such as deletions, insertions, translocations, and duplications of the CSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.,* Andreeff *et al.* (eds.), Introduction to Fluorescence *In Situ* Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999). The isolated nucleic acid molecules of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include a nucleic acid molecule of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level. Alternatively, detection techniques such as molecular beacons may be used, see Kostrikis *et al. Science* 279:1228-1229 (1998).

The isolated nucleic acid molecules of the present invention can be also be used as probes to detect, characterize, and quantify CSNA in, and isolate CSNA from, transcript-derived nucleic acid samples. In one embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A^+- selected RNA samples. In another embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.*, Schwarchzacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000). In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to CSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel

(1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000).

In another embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify and/or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In this embodiment, it is preferred that the probe or primer be derived from a nucleic acid molecule encoding a CSP. More preferably, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 101-194. Also preferred are probes or primers derived from a CSNA. More preferred are probes or primers derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-100.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); and McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995). Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR,

Eaton Publishing Company/Bio Techniques Books Division, 1998; and Siebert (ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995).

PCR and hybridization methods may be used to identify and/or isolate nucleic acid molecules of the present invention including allelic variants, homologous nucleic acid molecules and fragments. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules of the present invention that encode homologous proteins, analogs, fusion protein or muteins of the invention. Nucleic acid primers as described herein can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

These nucleic acid primers can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Pat. No. 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); international patent publications WO 97/19193 and WO 00/15779, and U.S. Patent Nos. 5,854,033 and 5,714,320, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene,

polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that comprise one or more of the nucleic acid molecules of the present invention.

In yet another embodiment, the invention is directed to single exon probes based on the CSNAs disclosed herein.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention provides vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acid molecules of the present invention in host cells (cloning vectors), for shuttling the nucleic acid molecules of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acid molecules of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acid molecules of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acid molecules of the present invention, alone or as fusion proteins with heterologous polypeptides (expression vectors). Vectors are by now well known in the art, and are described, *inter alia*, in Jones *et al.* (eds.),

Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*. Furthermore, a variety of vectors are available commercially. Use of existing vectors and modifications thereof are well within the skill in the art. Thus, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences that control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λGT10 and λGT11, and other phages, *e.g.*, M13 and filamentous single stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline,

chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as ura3-52, his3-D1, leu2-D1, trp1-D1 and lys2-201.

Insect cells may be chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda, e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

The host cells may also be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral

origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include, include but are not limited to, resistance to neomycin (G418), blasticidin, hygromycin and zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the nucleic acid molecules of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the

transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the trc promoter, a hybrid derived from the trp and lac promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, and the araBAD operon. Prokaryotic expression vectors may further include transcription terminators, such as the aspA terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1 promoter, the GAL1 promoter, the GAL10 promoter, ADH1 promoter, the promoters of the yeast α-mating system, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the CYC1 or ADH1 gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include, but are not limited to, those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 and the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the CSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β-globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows a high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the trc promoter, which is regulated by the lac operon, and the pL promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid Plac/ara-1 promoter and the PLtetO-1 promoter. The PLtetO-1 promoter takes advantage of the high expression levels from the PL promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the Tn10 tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one embodiment of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Such tags include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc.,

Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverley, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the polypeptides of the present invention can be expressed as a fusion to glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope, detectable by anti-HA antibody.

For secretion of expressed polypeptides, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful protein fusions include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusions for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins:

A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the α-agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well known in the art. *See* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999). A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g,* Cormack *et al.*, *Gene* 173: 33–38 (1996); U.S. Patent Nos. 6,090,919 and 5,804,387, the disclosures of which are incorporated herein by reference in their entireties) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g,*. Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene*

173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.,* Heim *et al., Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al., Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.,* Heikal *et al., Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patent Nos. 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. *See* also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999); Yang, *et al., J Biol Chem,* 273: 8212-6 (1998); Bevis *et al., Nature Biotechnology,* 20:83-7 (2002). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half-life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412, the disclosures of which are incorporated herein by reference in their entireties.

For long-term, high-yield recombinant production of the polypeptides of the present invention, stable expression is preferred. Stable expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo

Alto, CA, USA) allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid molecules of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed polypeptide in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation, and it is an aspect of the present invention to provide CSPs with such post-translational modifications.

In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid molecules of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid molecules of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed

from the nucleic acid molecules according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell (*e.g.*, conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda* — *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA) — *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2

cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (*e.g.*, from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well known and readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from colon are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human colon cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. *See, e.g.*, Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*.

Methods for introducing the vectors and nucleic acid molecules of the present invention into the host cells are well known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (*e.g.*, Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, *e.g.*, with $CaCl_2$, or a solution of Mg^{2+}, Mn^{2+}, Ca^{2+}, Rb^{+} or K^{+}, dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (*e.g.*, Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5α competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent E. coli Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent to take up exogenous DNA by electroporation by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided by BioRad (Richmond, CA, USA).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as a snail-gut extract, usually denoted Glusulase or Zymolyase, or an enzyme from *Arthrobacter luteus* to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+}. Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate to permeabilize the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with $CaPO_4$ or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for $CaPO_4$ transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found in, for example, ; Norton *et*

al. (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000). Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.*, Cheng *et al., Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al., Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

Purification of recombinantly expressed proteins is now well within the skill in the art and thus need not be detailed here. *See, e.g.*, Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001).

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tag, purification can be effected, at least in part, by means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides, including Fragments Muteins, Homologous Proteins, Allelic Variants, Analogs and Derivatives

Another aspect of the invention relates to polypeptides encoded by the nucleic acid molecules described herein. In a preferred embodiment, the polypeptide is a colon specific polypeptide (CSP). In an even more preferred embodiment, the polypeptide comprises an amino acid sequence of SEQ ID NO: 101-194 or is derived from a polypeptide having the amino acid sequence of SEQ ID NO: 101-194. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well known to those having ordinary skill in the art.

Polypeptides of the present invention may also comprise a part or fragment of a CSP. In a preferred embodiment, the fragment is derived from a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO: 101-194. Polypeptides of the present invention comprising a part or fragment of an entire CSP may or may not be CSPs. For example, a full-length polypeptide may be colon-specific, while a fragment thereof may be found in other tissues as well as in colon. A polypeptide that is not a CSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-CSP antibodies. In a preferred embodiment, the part or fragment is a CSP. Methods of determining whether a polypeptide of the present invention is a CSP are described *infra*.

Polypeptides of the present invention comprising fragments of at least 6 contiguous amino acids are also useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of a polypeptide of the present invention have utility in such a study.

Polypeptides of the present invention comprising fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize polypeptides of the present invention. *See, e.g.*, Lerner, Nature 299: 592-596 (1982); Shinnick *et al.*, *Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al.*, *Science* 219: 660-6 (1983). As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic and are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the polypeptides of the present invention have utility as immunogens.

Polypeptides comprising fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire polypeptide, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind

specifically to the polypeptide of interest. See U.S. Patent Nos. 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

The polypeptide of the present invention thus preferably is at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the polypeptide of the present invention is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger polypeptides having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments by truncating the nucleic acid molecule, *e.g.*, a CSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally occurring polypeptide. Methods of producing polypeptide fragments are well known in the art. *See, e.g.,* Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra.* In one embodiment, a polypeptide comprising only a fragment, preferably a fragment of a CSP, may be produced by chemical or enzymatic cleavage of a CSP polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule of the present invention encoding a fragment, preferably of a CSP, in a host cell.

Polypeptides of the present invention are also inclusive of mutants, fusion proteins, homologous proteins and allelic variants.

A mutant protein, or mutein, may have the same or different properties compared to a naturally occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native polypeptide. Small deletions and insertions can often be found that do not alter the function of a protein. Muteins may or may not be colon-specific. Preferably, the mutein is colon-specific. More preferably the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 101-194. Accordingly, in a preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a CSP comprising an amino acid sequence of

SEQ ID NO: 101-194. In a yet more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 101-194.

A mutein may be produced by isolation from a naturally occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein is produced from a host cell comprising a mutated nucleic acid molecule compared to the naturally occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid molecule of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is colon-specific, as described below. Multiple random mutations can be introduced into the gene by methods well known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well known in the art. *See, e.g.*, Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), as well as U.S. Patent No. 5,223,408, which is herein incorporated by reference in its entirety.

The invention also contemplates polypeptides that are homologous to a polypeptide of the invention. In a preferred embodiment, the polypeptide is homologous to a CSP. In an even more preferred embodiment, the polypeptide is homologous to a CSP selected from the group having an amino acid sequence of SEQ ID NO: 101-194. By homologous polypeptide it is means one that exhibits significant sequence identity to a CSP, preferably a CSP having an amino acid sequence of SEQ ID NO: 101-194. By significant sequence identity it is meant that the homologous polypeptide exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more

preferably at least 70%, yet more preferably at least 80% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 101-194. More preferred are homologous polypeptides exhibiting at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 101-194. Most preferably, the homologous polypeptide exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 101-194. In a preferred embodiment, the amino acid substitutions of the homologous polypeptide are conservative amino acid substitutions as discussed above.

Homologous polypeptides of the present invention also comprise polypeptide encoded by a nucleic acid molecule that selectively hybridizes to a CSNA or an antisense sequence thereof. In this embodiment, it is preferred that the homologous polypeptide be encoded by a nucleic acid molecule that hybridizes to a CSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. More preferred is a homologous polypeptide encoded by a nucleic acid sequence which hybridizes to a CSNA selected from the group consisting of SEQ ID NO: 1-100 or a homologous polypeptide encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a CSP, preferably an CSP of SEQ ID NO: 101-194 under low stringency, moderate stringency or high stringency conditions, as defined herein.

Homologous polypeptides of the present invention may be naturally occurring and derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, or baboon, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 101-194. The homologous polypeptide may also be a naturally occurring polypeptide from a human, when the CSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally occurring homologous polypeptide may be isolated and used to

express the homologous polypeptide recombinantly. The homologous polypeptide may also be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. Alternatively, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a CSP. In a preferred embodiment, the homologous polypeptide encodes a polypeptide that is a CSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated polpeptide not only identical in sequence to those described with particularity herein, but also to provide isolated polypeptide ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, polypeptides of the present invention are also inclusive of those encoded by an allelic variant of a nucleic acid molecule encoding a CSP. In this embodiment, it is preferred that the polypeptide be encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 101-194. More preferred is that the polypeptide be encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1-100.

Polypeptides of the present invention are also inclusive of derivative polypeptides encoded by a nucleic acid molecule according to the instant invention. In this embodiment, it is preferred that the polypeptide be a CSP. Also preferred are derivative polypeptides having an amino acid sequence selected from the group consisting of SEQ ID NO: 101-194 and which has been acetylated, carboxylated, phosphorylated, glycosylated, ubiquitinated or other PTMs. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H. In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed November 11, 2002), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications include, but are not limited to: (Z)-dehydrobutyrine; 1-chondroitin sulfate-L-aspartic acid ester; 1'-glycosyl-L-tryptophan; 1'-phospho-L-histidine; 1-thioglycine; 2'-(S-L-cysteinyl)-L-histidine; 2'-[3-carboxamido (trimethylammonio)propyl]-L-histidine; 2'-alpha-mannosyl-L-tryptophan; 2-methyl-L-glutamine; 2-oxobutanoic acid; 2-pyrrolidone carboxylic acid; 3'-(1'-L-histidyl)-L-tyrosine; 3'-(8alpha-FAD)-L-histidine; 3'-(S-L-cysteinyl)-L-tyrosine; 3', 3",5'-triiodo-L-thyronine; 3'-4'-phospho-L-tyrosine; 3-hydroxy-L-proline; 3'-methyl-L-histidine; 3-methyl-L-lanthionine; 3'-phospho-L-histidine; 4'-(L-tryptophan)-L-tryptophyl quinone; 42 N-cysteinyl-glycosylphosphatidylinositolethanolamine; 43 -(T-L-histidyl)-L-tyrosine; 4-hydroxy-L-arginine; 4-hydroxy-L-lysine; 4-hydroxy-L-proline; 5'-(N6-L-lysine)-L-topaquinone; 5-hydroxy-L-lysine; 5-methyl-L-arginine; alpha-1-microglobulin-Ig alpha

complex chromophore; bis-L-cysteinyl bis-L-histidino diiron disulfide; bis-L--cysteinyl-L-N3'-histidino-L-serinyl tetrairon' tetrasulfide; chondroitin sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; D-alanine; D-allo-isoleucine; D-asparagine; dehydroalanine; dehydrotyrosine; dermatan 4-sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; D-glucuronyl-N-glycine; dipyrrolylmethanemethyl-L-cysteine; D-leucine; D-methionine; D-phenylalanine; D-serine; D-tryptophan; glycine amide; glycine oxazolecarboxylic acid; glycine thiazolecarboxylic acid; heme P450-bis-L-cysteine-L-tyrosine; heme-bis-L-cysteine; hemediol-L-aspartyl ester-L-glutamyl ester; hemediol-L-aspartyl ester-L-glutamyl ester-L-methionine sulfonium; heme-L-cysteine; heme-L-histidine; heparan sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; heme P450-bis-L-cysteine-L-lysine; hexakis-L-cysteinyl hexairon hexasulfide; keratan sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-threonine; L oxoalanine- lactic acid; L phenyllactic acid; l'-(8alpha-FAD)-L-histidine; L-2'.4',5'-topaquinone; L-3',4'-dihydroxyphenylalanine; L-3'.4'.5'-trihydroxyphenylalanine; L-4'-bromophenylalanine; L-6'-bromotryptophan; L-alanine amide; L-alanyl imidazolinone glycine; L-allysine; L-arginine amide; L-asparagine amide; L-aspartic 4-phosphoric anhydride; L-aspartic acid 1-amide; L-beta-methylthioaspartic acid; L-bromohistidine; L-citrulline; L-cysteine amide; L-cysteine glutathione disulfide; L-cysteine methyl disulfide; L-cysteine methyl ester; L-cysteine oxazolecarboxylic acid; L-cysteine oxazolinecarboxylic acid; L-cysteine persulfide; L-cysteine sulfenic acid; L-cysteine sulfinic acid; L-cysteine thiazolecarboxylic acid; L-cysteinyl homocitryl molybdenum-heptairon-nonasulfide; L-cysteinyl imidazolinone glycine; L-cysteinyl molybdopterin; L-cysteinyl molybdopterin guanine dinucleotide; L-cystine; L-erythro-beta-hydroxyasparagine; L-erythro-beta-hydroxyaspartic acid; L-gamma-carboxyglutamic acid; L-glutamic acid 1-amide; L-glutamic acid 5-methyl ester; L-glutamine amide; L-glutamyl 5-glycerylphosphorylethanolamine; L-histidine amide; L-isoglutamyl-polyglutamic acid; L-isoglutamyl-polyglycine; L-isoleucine amide; L-lanthionine; L-leucine amide; L-lysine amide; L-lysine thiazolecarboxylic acid; L-lysinoalanine; L-methionine amide; L-methionine sulfone; L-phenyalanine thiazolecarboxylic acid; L-phenylalanine amide; L-proline amide; L-selenocysteine; L-selenocysteinyl molybdopterin guanine dinucleotide; L-serine amide; L-serine thiazolecarboxylic acid; L-seryl imidazolinone glycine; L-T-bromophenylalanine; L-T-bromophenylalanine; L-threonine amide; L-thyroxine; L-tryptophan amide; L-tryptophyl quinone; L-tyrosine amide; L-valine amide; meso-

lanthionine; N-(L-glutamyl)-L-tyrosine; N-(L-isoaspartyl)-glycine; N-(L-isoaspartyl)-L-cysteine; N,N,N-trimethyl-L-alanine; N,N-dimethyl-L-proline; N2-acetyl-L-lysine; N2-succinyl-L-tryptophan; N4-(ADP-ribosyl)-L-asparagine; N4-glycosyl-L-asparagine; N4-hydroxymethyl-L-asparagine; N4-methyl-L-asparagine; N5-methyl-L-glutamine; N6- 1 -carboxyethyl-L-lysine; N6-(4-amino hydroxybutyl)-L-lysine; N6-(L-isoglutamyl)-L-lysine; N6-(phospho-5'-adenosine)-L-lysine; N6-(phospho-5'-guanosine)-L-tysine; N6,N6,N6-trimethyl-L-lysine; N6,N6-dimethyl-L-lysine; N6-acetyl-L-lysine; N6-biotinyl-L-lysine; N6-carboxy-L-lysine; N6-formyl-L-lysine; N6-glycyl-L-lysine; N6-lipoyl-L-lysine; N6-methyl-L-lysine; N6-methyl-N6-poly(N-methyl-propylamine)-L-lysine; N6-mureinyl-L-lysine; N6-myristoyl-L-lysine; N6-palmitoyl-L-lysine; N6-pyridoxal phosphate-L-lysine; N6-pyruvic acid 2-iminyl-L-lysine; N6-retinal-L-lysine; N-acetylglycine; N-acetyl-L-glutamine; N-acetyl-L-alanine; N-acetyl-L-aspartic acid; N-acetyl-L-cysteine; N-acetyl-L-glutamic acid; N-acetyl-L-isoleucine; N-acetyl-L-methionine; N-acetyl-L-proline; N-acetyl-L-serine; N-acetyl-L-threonine; N-acetyl-L-tyrosine; N-acetyl-L-valine; N-alanyl-glycosylphosphatidylinositolethanolamine; N-asparaginyl-glycosylphosphatidylinositolethanolarnine; N-aspartyl-glycosylphosphatidylinositolethanolamine; N-formylglycine; N-formyl-L-methionine; N-glycyl-glycosylphosphatidylinositolethanolamine; N-L-glutamyl-poly-L-glutamic acid; N-methylglycine; N-methyl-L-alanine; N-methyl-L-methionine; N-methyl-L-phenylalanine; N-myristoyl-glycine; N-palmitoyl-L-cysteine; N-pyruvic acid 2-iminyl-L-cysteine; N-pyruvic acid 2-iminyl-L-valine; N-seryl-glycosylphosphatidylinositolethanolamine; N-seryl-glycosyCSPhingolipidinositolethanolamine; O-(ADP-ribosyl)-L-serine; O-(phospho-5'-adenosine)-L-threonine; O-(phospho-5'-DNA)-L-serine; O-(phospho-5'-DNA)-L-threonine; O-(phospho-5'rRNA)-L-serine; O-(phosphoribosyl dephospho-coenzyme A)-L-serine; O-(sn-l-glycerophosphoryl)-L-serine; O4'-(8alpha-FAD)-L-tyrosine; O4'-(phospho-5'-adenosine)-L-tyrosine; O4'-(phospho-5'-DNA)-L-tyrosine; O4'-(phospho-5'-RNA)-L-tyrosine; O4'-(phospho-5'-uridine)-L-tyrosine; O4-glycosyl-L-hydroxyproline; O4'-glycosyl-L-tyrosine; O4'-sulfo-L-tyrosine; O5-glycosyl-L-hydroxylysine; O-glycosyl-L-serine; O-glycosyl-L-threonine; omega-N-(ADP-ribosyl)-L-arginine; omega-N-omega-N'-dimethyl-L-arginine; omega-N-methyl-L-arginine; omega-N-omega-N-dimethyl-L-arginine; omega-N-phospho-L-arginine; O'octanoyl-L-serine; O-palmitoyl-L-serine; O-palmitoyl-L-threonine; O-phospho-L-serine; O-phospho-L-threonine; O-phosphopantetheine-L-serine; phycoerythrobilin-bis-L-cysteine; phycourobilin-bis-L-

cysteine; pyrroloquinoline quinone; pyruvic acid; S hydroxycinnamyl-L-cysteine; S-(2-aminovinyl) methyl-D-eysteine; S-(2-aminovinyl)-D-cysteine; S-(6-FW-L-cysteine; S-(8alpha-FAD)-L-cysteine; S-(ADP-ribosyl)-L-cysteine; S-(L-isoglutamyl)-L-cysteine; S-12-hydroxyfarnesyl-L-cysteine; S-acetyl-L-cysteine; S-diacylglycerol-L-cysteine; S-diphytanylglycerot diether-L-cysteine; S-farnesyl-L-cysteine; S-geranylgeranyl-L-cysteine; S-glycosyl-L-cysteine; S-glycyl-L-cysteine; S-methyl-L-cysteine; S-nitrosyl-L-cysteine; S-palmitoyl-L-cysteine; S-phospho-L-cysteine; S-phycobiliviolin-L-cysteine; S-phycocyanobilin-L-cysteine; S-phycoerythrobilin-L-cysteine; S-phytochromobilin-L-cysteine; S-selenyl-L-cysteine; S-sulfo-L-cysteine; tetrakis-L-cysteinyl diiron disulfide; tetrakis-L-cysteinyl iron; tetrakis-L-cysteinyl tetrairon tetrasulfide; trans-2,3-cis 4-dihydroxy-L-proline; tris-L-cysteinyl triiron tetrasulfide; tris-L-cysteinyl triiron trisulfide; tris-L-cysteinyl-L-aspartato tetrairon tetrasulfide; tris-L-cysteinyl-L-cysteine persulfido-bis-L-glutamato-L-histidino tetrairon disulfide trioxide; tris-L-cysteinyl-L-N3'-histidino tetrairon tetrasulfide; tris-L-cysteinyl-L-Nl'-histidino tetrairon tetrasulfide; and tris-L-cysteinyl-L-serinyl tetrairon tetrasulfide.

Additional examples of PTMs may be found in web sites such as the Delta Mass database based on Krishna, R. G. and F. Wold (1998). Posttranslational Modifications. Proteins - Analysis and Design. R. H. Angeletti. San Diego, Academic Press. 1: 121-206. ; Methods in Enzymology, 193, J.A. McClosky (ed) (1990), pages 647-660; Methods in Protein Sequence Analysis edited by Kazutomo Imahori and Fumio Sakiyama, Plenum Press, (1993) "Post-translational modifications of proteins" R.G. Krishna and F. Wold pages 167-172; "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. Nucleic Acids Res. 29; 332-335 (2001) "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. Nucleic Acids Research, 27: 370-372 (1999); and "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al.Nucleic Acids Res 27(1):237-239 (1999) see also,WO 02/21139A2, the disclosure of which is incorporated herein by reference in its entirety.

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known.

One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signalling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell

compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g, p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally

modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule may also be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568,

Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

Polypeptides of the present invention, including full length polypeptides, fragments and fusion proteins, can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to polypeptides of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

Polypeptides of the present invention, including full length polypeptide, fragments and fusion proteins, can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-CSP antibodies.

Polypeptides of the present invention, including full length polypeptide, fragments and fusion proteins, can also usefully be conjugated to polyethylene glycol (PEG);

PEGylation increases the serum half life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999). PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

Polypeptides of the present invention are also inclusive of analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, this polypeptide is a CSP. In a more preferred embodiment, this polypeptide is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 101-194. Also preferred is an analog polypeptide comprising one or more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally occurring polypeptide. In one embodiment, the analog is structurally similar to a CSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --$CH(OH)CH_2$-- and --CH_2SO--. In another embodiment, the analog comprises substitution of one or more amino acids of a CSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g.*, Kole *et al.*, *Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.

Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, *inter alia*, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones,

Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993).

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinoyl--(9-fluorenylmethoxycarbonyl)-L-lysine (FMOC biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The FMOC and *t*BOC derivatives of dabcyl-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyl chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyl quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS--FMOC-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated FMOC synthesis of peptides using (FMOC)--TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other FMOC-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-

2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)-∃-alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

Another aspect of the present invention relates to the fusion of a polypeptide of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide of the present invention is a CSP. In a more preferred embodiment, the polypeptide of the present invention that is fused to a heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 101-194, or is a mutein,

homologous polypeptide, analog or derivative thereof. In an even more preferred embodiment, the fusion protein is encoded by a nucleic acid molecule comprising all or part of the nucleic acid sequence of SEQ ID NO: 1-100, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-100.

The fusion proteins of the present invention will include at least one fragment of a polypeptide of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the polypeptide of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of a polypeptide of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and preferably at least 15, 20, or 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particularly useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g.*, Ausubel, Chapter 16, (1992), *supra*. Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins into the periplasmic space or extracellular milieu for prokaryotic hosts or into the culture medium for eukaryotic cells through incorporation of secretion signals and/or leader sequences. For example, a His^6 tagged protein can be

purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful fusion proteins of the present invention include those that permit use of the polypeptide of the present invention as bait in a yeast two-hybrid system. *See* Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); Colas *et al.*, *Nature* 380, 548-550 (1996); Norman, T. *et al.*, *Science* 285, 591-595 (1999); Fabbrizio *et al.*, *Oncogene* 18, 4357-4363 (1999); Xu *et al.*, *Proc Natl Acad Sci U S A.* 94, 12473-12478 (1997); Yang, *et al.*, *Nuc. Acids Res.* 23, 1152-1156 (1995); Kolonin *et al.*, *Proc Natl Acad Sci U S A* 95, 14266-14271 (1998); Cohen *et al.*, *Proc Natl Acad Sci U S A* 95, 14272-14277 (1998); Uetz, *et al. Nature* 403, 623-627(2000); Ito, *et al.*, *Proc Natl Acad Sci U S A* 98, 4569-4574 (2001). Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded polypeptide on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above.

The polypeptides of the present invention can also usefully be fused to protein toxins, such as Pseudomonas exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins, β-galactosidase, biotin trpE, protein A, β-lactamase, α-amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast α mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. *See, e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well known in the art (*e.g.*, a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the CSP.

As further described below, the polypeptides of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize polypeptides of the present invention including CSPs and their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly CSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of CSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of CSPs.

One may determine whether polypeptides of the present invention including CSPs, muteins, homologous proteins or allelic variants or fusion proteins of the present invention are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the polypeptide at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992);

combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, (Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides or fusion proteins of the present invention is well known and within the skill of one having ordinary skill in the art. *See, e.g.,* Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated polypeptides or fusion proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous and non-proteinaceous material and are well known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated polypeptide or fusion protein of the present invention are used as therapeutic agents, such as in vaccines and replacement therapy, the isolated polypeptides of the present invention are also useful at lower purity. For example, partially purified polypeptides of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In a preferred embodiment, the purified and substantially purified polypeptides of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides or fusion proteins of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent. For example, the peptides of the invention may be stabilized by covalent linkage to albumin. See, U.S. Patent No. 5,876,969, the contents of which are hereby incorporated in its entirety.

For example, the polypeptides or fusion proteins of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the polypeptides or fusion proteins of the present invention can be used

to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized polypeptide or fusion protein of the present invention.

As another example, the polypeptides or fusion proteins of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides and fusion proteins of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the polypeptide or fusion protein of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound polypeptide or fusion protein to indicate biologic interaction there between. The polypeptides or fusion proteins of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the polypeptide or fusion protein of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound polypeptide or fusion protein to indicate biological interaction there between.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention. In a preferred embodiment, the antibodies are specific for a polypeptide that is a CSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 101-194, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also

due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a CSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or vis versa. In addition, alternative splice forms of a CSP may be indicative of cancer. Differential degradation of the C or N-terminus of a CSP may also be a marker or target for anticancer therapy. For example, an CSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-CSP polypeptides by at least two-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the polypeptide of the present invention in samples derived from human colon.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to 1×10^{-13} M proving especially useful.

The antibodies of the present invention can be naturally occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In such case, antibodies to the polypeptides of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the polypeptide of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic

mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patent Nos. 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention are also usefully obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster), lagomorphs (typically rabbits), and also larger mammals, such as sheep, goats, cows, and horses; or egg laying birds or reptiles such as chickens or alligators. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the polypeptide of the present invention. One form of avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000.

As discussed above, virtually all fragments of 8 or more contiguous amino acids of a polypeptide of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide of the present invention to other moieties. For example, polypeptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. *See* Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996). Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the polypeptides of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the polypeptides of the present invention. Antibodies from avian species may have particular advantage in detection of the polypeptides of the present invention, in human serum or tissues (Vikinge et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998). Following immunization, the antibodies of the present invention can be obtained using any art-accepted technique. Such techniques are well known in the art and are described in detail in references such as Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); and Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997).

Briefly, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the polypeptides of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the polypeptides of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further

described in U.S. Patent No. 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant antibody production of whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.*, Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*, 4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994). Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.*, Barbas (2001), *supra*; Kay, *supra*; and Abelson, *supra*.

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell. Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention. For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.*, Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, J. Biotechnol. 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); and Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998).

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.*, Li *et al.*, *Protein Expr. Purif.* 21(1): 121-8

(2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992).

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995).

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock et al., *J. Immunol Methods.* 231: 147-57 (1999); Young et al., *Res. Immunol.* 149: 609-10 (1998); and Limonta et al., *Immunotechnology* 1: 107-13 (1995).

Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells. Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998) review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies. Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem.* (Tokyo) 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999).

The invention further provides antibody fragments that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention. Among such useful fragments are Fab, Fab', Fv, $F(ab)'_2$, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

The present invention also relates to antibody derivatives that bind specifically to one or more of the polypeptides of the present invention, to one or more of the

polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus are more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful method is PEGylation to increase the serum half life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, Morrison *et al.*, *Proc. Natl. Acad. Sci USA*.81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985); and U.S. Patent No. 5,807,715 the disclosure of which is incorporated herein by reference in its entirety. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); and U.S. Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties. Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. Accordingly, the present invention includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be

transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention. The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label can usefully be an enzyme that catalyzes production and local deposition of a detectable product. Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well known, and include alkaline phosphatase, β-galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopryanoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and

requiring only small amounts of antibody. *See, e.g.*, Thorpe *et al.*, *Methods Enzymol.* 133: 331-53 (1986); Kricka *et al.*, *J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995). Kits for such enhanced chemiluminescent detection (ECL) are available commercially. The antibodies can also be labeled using colloidal gold.

As another example, when the antibodies of the present invention are used, *e.g.*, for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores. There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention. For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention. For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P, ^{32}P, ^{35}S, ^{3}H, and ^{125}I. As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th, ^{227}Ac, ^{225}Ac, ^{223}Ra, ^{213}Bi, ^{212}Pb, ^{212}Bi, ^{211}At, ^{203}Pb, ^{194}Os, ^{188}Re, ^{186}Re, ^{153}Sm, ^{149}Tb, ^{131}I, ^{125}I, ^{111}In, ^{105}Rh, ^{99m}Tc, ^{97}Ru, ^{90}Y, ^{90}Sr, ^{88}Y, ^{72}Se, ^{67}Cu, or ^{47}Sc.

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application as for which they were mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the polypeptides of the present invention. Commonly, the antibody in such immunotoxins is conjugated to Pseudomonas exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, or ricin. *See* Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998).

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, attached to a substrate. Substrates can be porous or nonporous, planar or nonplanar. For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography. For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microsphere can then be used for isolation of cells that express or display the polypeptides of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the CSPs of the present invention or to polypeptides encoded by the CSNAs of the invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding a CSP. In a preferred embodiment, the CSP comprises an amino acid sequence selected from SEQ ID NO: 101-194, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a CSNA of the invention, preferably a CSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-100, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human CSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well known in the art. *See, e.g.*, Hogan *et al.*, Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.*, Paterson

et al., *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and U.S. Patent No. 4,873,191, herein incorporated by reference in its entirety); retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g.,* Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g.,* Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g.,* Lo, 1983, *Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g.,* Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g.,* Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.,* Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)).

The present invention provides for transgenic animals that carry the transgene (*i.e.*, a nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i.e. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, *e.g.*, the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232- 6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g.*, Gu *et al.*, *Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g.*, Smithies *et al.*, *Nature* 317: 230-234 (1985); Thomas *et al.*, *Cell* 51: 503-512 (1987); Thompson *et al.*, *Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention in vivo. In another embodiment, techniques known in the art are used to generate knockouts in cells that

contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g.,* Thomas, *supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site in vivo using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a patient in vivo. Such cells may be obtained from an animal or patient or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered in vitro using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.,* U.S. Patent Nos. 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the

development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention is a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 101-194 and SEQ ID NO: 1-100 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises

the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid

of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence. In addition, the invention includes a method of using patterns of expression associated with either the nucleic acids or proteins in a computer-based method to diagnose disease.

Diagnostic Methods for colon Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by comparing expression of a CSNA or a CSP in a human patient that has or may have colon cancer, or who is at risk of developing colon cancer, with the expression of a CSNA or a CSP in a normal human control. For purposes of the present invention, "expression of a CSNA" or "CSNA expression" means the quantity of CSNA mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of a CSP" or "CSP expression" means the amount of CSP that can be measured by any method known in the art or the level of translation of a CSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing colon cancer in a patient, in particular adenocarcinoma, by analyzing for changes in levels of CSNA or CSP in cells, tissues, organs or bodily fluids compared with levels of CSNA or CSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of a CSNA or CSP in the patient versus the normal human control is associated with the presence of colon cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing colon cancer in a patient by analyzing changes in the structure of the mRNA of a CSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing colon cancer in a patient by analyzing changes in a CSP compared to a CSP from a normal patient. These changes include, *e.g.*, alterations, including post translational modifications such as glycosylation and/or phosphorylation of the CSP or changes in the subcellular CSP localization.

For purposes of the present invention, diagnosing means that CSNA or CSP levels are used to determine the presence or absence of disease in a patient. As will be understood by those of skill in the art, measurement of other diagnostic parameters may be required for definitive diagnosis or determination of the appropriate treatment for the disease. The determination may be made by a clinician, a doctor, a testing laboratory, or a patient using an over the counter test. The patient may have symptoms of disease or may be asymptomatic. In addition, the CSNA or CSP levels of the present invention may be used as screening marker to determine whether further tests or biopsies are warranted. In addition, the CSNA or CSP levels may be used to determine the vulnerability or susceptibility to disease.

In a preferred embodiment, the expression of a CSNA is measured by determining the amount of a mRNA that encodes an amino acid sequence selected from SEQ ID NO: 101-194, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the CSNA expression that is measured is the level of expression of a CSNA mRNA selected from SEQ ID NO: 1-100, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acid molecules. CSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. CSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of a CSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, CSNA expression may be compared to a known control, such as normal colon nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of a CSP is measured by determining the level of a CSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 101-194, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression

of a CSNA or CSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of colon cancer. The expression level of a CSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the CSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See*, e.g, Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the CSP structure may be determined by any method known in the art, including, *e.g.*, using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to a CSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-CSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the CSP will bind to the anti-CSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-CSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the CSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of an CSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure CSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-CSP antibody is attached to a solid support and an allocated amount of a labeled CSP and a sample of interest are incubated with the solid support. The amount of labeled CSP attached to the solid support can be correlated to the quantity of a CSP in the sample.

Of the proteomic approaches, 2D PAGE is a well known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of a CSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more CSNAs of interest. In this approach, all or a portion of one or more CSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy

material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. As used herein "blood" includes whole blood, plasma, serum, circulating epithelial cells, constituents, or any derivative of blood.

In addition to detection in bodily fluids, the proteins and nucleic acids of the invention are suitable to detection by cell capture technology. Whole cells may be captured by a variety methods for example magnetic separation, U.S. Patent. Nos. 5,200,084; 5,186,827; 5,108,933; 4,925,788, the disclosures of which are incorporated herein by reference in their entireties. Epithelial cells may be captured using such products as Dynabeads® or CELLection™ (Dynal Biotech, Oslo, Norway). Alternatively, fractions of blood may be captured, e.g., the buffy coat fraction (50mm cells isolated from 5ml of blood) containing epithelial cells. In addition, cancer cells may be captured using the techniques described in WO 00/47998, the disclosure of which is incorporated herein by reference in its entirety. Once the cells are captured or concentrated, the proteins or nucleic acids are detected by the means described in the subject application. Alternatively, nucleic acids may be captured directly from blood samples, see U.S. Patent Nos. 6,156,504, 5,501,963; or WO 01/42504 , the disclosures of which are incorporated herein by reference in their entireties.

In a preferred embodiment, the specimen tested for expression of CSNA or CSP includes without limitation colon tissue, fecal samples, colonocytes, colon cells grown in cell culture, blood, serum, lymph node tissue, and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary colon cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, lungs, and adrenal glands. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediatinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration.

Colonocytes represent an important source of the CSP or CSNAs because they provide a picture of the immediate past metabolic history of the GI tract of a subject. In addition, such cells are representative of the cell population from a statistically large sampling frame reflecting the state of the colonic mucosa along the entire length of the colon in a non-invasive manner, in contrast to a limited sampling by colonic biopsy using an invasive procedure involving endoscopy. Specific examples of patents describing the isolatation colonocytes include U.S. Patent Nos. 6,335,193; 6,020,137 5,741,650;

6,258,541; US 2001 0026925 A1; WO 00/63358 A1, the disclosures of which are incorporated herein by reference in their entireties.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of a CSNA or CSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other CSNA or CSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular CSNA or CSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more CSNA and/or CSP in a sample from a patient suspected of having colon cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a CSNA and/or CSP and then ascertaining whether the patient has colon cancer from the expression level of the CSNA or CSP. In general, if high expression relative to a control of a CSNA or CSP is indicative of colon cancer, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least one and a half times higher, and more preferably are at least two times higher, still more preferably five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is indicative of colon cancer, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least one and a half times lower, and more preferably are at least two times lower, still more preferably five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether colon cancer has metastasized in a patient. One may identify whether the colon cancer has metastasized by measuring the expression levels and/or structural alterations of one or more CSNAs and/or CSPs in a variety of tissues. The presence of a CSNA or CSP in a certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of a CSNA or CSP is associated with colon cancer. Similarly, the presence of a CSNA or CSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a CSNA or CSP is associated with colon cancer. Further, the presence of a structurally altered CSNA or CSP that is associated with colon cancer is also indicative of metastasis.

In general, if high expression relative to a control of a CSNA or CSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the CSNA or CSP is at least one and a half times higher, and more preferably are at least two times higher, still more preferably five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the CSNA or CSP is at least one and a half times lower, and more preferably are at least two times lower, still more preferably five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

Staging

The invention also provides a method of staging colon cancer in a human patient. The method comprises identifying a human patient having colon cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more CSNAs or CSPs. First, one or more tumors from a variety of patients are staged according to procedures well known in the art, and the expression levels of one or more CSNAs or CSPs is determined for each stage to obtain a standard expression level for each CSNA and CSP. Then, the CSNA or CSP expression levels of the CSNA or CSP are determined in a biological sample from a patient whose stage of cancer is not known. The CSNA or CSP expression levels from the patient are

then compared to the standard expression level. By comparing the expression level of the CSNAs and CSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a CSNA or CSP to determine the stage of a colon cancer.

Monitoring

Further provided is a method of monitoring colon cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the colon cancer. The monitoring may determine if there has been a reoccurrence and, if so, determine its nature. The method comprises identifying a human patient that one wants to monitor for colon cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more CSNAs or CSPs, and comparing the CSNA or CSP levels over time to those CSNA or CSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a CSNA or CSP that are associated with colon cancer.

If increased expression of a CSNA or CSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of a CSNA or CSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a CSNA or CSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting a decrease in the expression level of a CSNA or CSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of CSNAs or CSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of colon cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a CSNA and/or CSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more CSNAs and/or CSPs are detected. The presence of higher (or lower) CSNA or CSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly colon cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more CSNAs and/or CSPs of the invention can also be monitored by analyzing levels of expression of the CSNAs and/or CSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in a CSG, thereby determining if a human with the genetic lesion is susceptible to developing colon cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing colon cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the CSGs of this invention, a chromosomal rearrangement of a CSG, an aberrant modification of a CSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a CSG. Methods to detect such lesions in the CSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

Methods of Detecting Noncancerous colon Diseases

The present invention also provides methods for determining the expression levels and/or structural alterations of one or more CSNAs and/or CSPs in a sample from a patient suspected of having or known to have a noncancerous colon disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of a CSNA and/or CSP, comparing the expression level or structural alteration of the CSNA or CSP to a normal colon control, and then

ascertaining whether the patient has a noncancerous colon disease. In general, if high expression relative to a control of a CSNA or CSP is indicative of a particular noncancerous colon disease, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is indicative of a noncancerous colon disease, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether a CSNA and/or CSP is associated with a particular noncancerous colon disease by obtaining colon tissue from a patient having a noncancerous colon disease of interest and determining which CSNAs and/or CSPs are expressed in the tissue at either a higher or a lower level than in normal colon tissue. In another embodiment, one may determine whether a CSNA or CSP exhibits structural alterations in a particular noncancerous colon disease state by obtaining colon tissue from a patient having a noncancerous colon disease of interest and determining the structural alterations in one or more CSNAs and/or CSPs relative to normal colon tissue.

Methods for Identifying colon Tissue

In another aspect, the invention provides methods for identifying colon tissue. These methods are particularly useful in, *e.g.*, forensic science, colon cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is colon tissue or has colon tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising colon tissue or having colon tissue-like characteristics, determining whether the sample expresses one or more CSNAs and/or CSPs, and, if the sample expresses one or more CSNAs and/or CSPs, concluding that the sample comprises colon tissue. In a preferred embodiment, the CSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 101-194, or a

homolog, allelic variant or fragment thereof. In a more preferred embodiment, the CSNA has a nucleotide sequence selected from SEQ ID NO: 1-100, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses a CSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether a CSP is expressed. Determining whether a sample expresses a CSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the CSP has an amino acid sequence selected from SEQ ID NO: 101-194, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two CSNAs and/or CSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five CSNAs and/or CSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is colon tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into colon tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new colon tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

Methods for Producing and Modifying colon Tissue

In another aspect, the invention provides methods for producing engineered colon tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing a CSNA or a CSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of colon tissue cells. In a preferred embodiment, the cells are pleuripotent. As is well known in the art, normal colon tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered colon tissue or cells comprises one of these cell types. In another embodiment,

the engineered colon tissue or cells comprises more than one colon cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the colon cell tissue. Methods for manipulating culture conditions are well known in the art.

Nucleic acid molecules encoding one or more CSPs are introduced into cells, preferably pleuripotent cells. In a preferred embodiment, the nucleic acid molecules encode CSPs having amino acid sequences selected from SEQ ID NO: 101-194, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1-100, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, a CSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well known in the art and are described in detail, *supra*.

Artificial colon tissue may be used to treat patients who have lost some or all of their colon function.

Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, fusion proteins, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, or inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises a CSNA or part thereof. In a more preferred embodiment, the CSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-100, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises a CSP or fragment thereof. In a more preferred embodiment, the pharmaceutical composition comprises a CSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 101-194, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-CSP antibody, preferably an antibody that specifically binds to a CSP having an amino acid that is selected from the group consisting of SEQ ID NO: 101-194, or an antibody that binds to a polypeptide that

is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art that is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000) and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, cornstarch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically. For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid

ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example CSP polypeptide, fusion protein, or fragments thereof, antibodies specific for CSP, agonists, antagonists or inhibitors of CSP, which ameliorates the signs or symptoms of the disease or prevent progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed

by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention, *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of colon function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for the purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further described in U.S. Patent Nos. 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g.*, Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid molecule of the present invention is administered. The nucleic acid molecule can be delivered in a vector that drives expression of a CSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a CSP are administered, for example, to complement a deficiency in the native CSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses,

adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g.*, Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid molecule encodes a CSP having the amino acid sequence of SEQ ID NO: 101-194, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a CSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in CSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a CSP having the amino acid sequence of SEQ ID NO: 101-194, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of a CSG antisense nucleic acid, are administered to downregulate transcription and/or translation of a CSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a CSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to CSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995).

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the CSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); and McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000). Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a CSP, preferably a CSP comprising an amino acid sequence of SEQ ID NO: 101-194, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-100, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a CSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant CSP defect.

Protein compositions are administered, for example, to complement a deficiency in native CSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to CSP. The immune response can be used to modulate activity of CSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate CSP.

In a preferred embodiment, the polypeptide administered is a CSP comprising an amino acid sequence of SEQ ID NO: 101-194, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-100, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well known, antibody compositions are administered, for example, to antagonize activity of CSP, or to target therapeutic agents to sites of CSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a CSP comprising an amino acid sequence of SEQ ID NO: 101-194, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to a

CSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-100, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to a CSP or have a modulatory effect on the expression or activity of a CSP. Modulators which decrease the expression or activity of CSP (antagonists) are believed to be useful in treating colon cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of a CSP can also be designed, synthesized and tested for use in the imaging and treatment of colon cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the CSPs identified herein. Molecules identified in the library as being capable of binding to a CSP are key candidates for further evaluation for use in the treatment of colon cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a CSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of CSP is administered. Antagonists of CSP can be produced using methods generally known in the art. In particular, purified CSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a CSP.

In other embodiments a pharmaceutical composition comprising an agonist of a CSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a CSP comprising an amino acid sequence of SEQ ID NO: 101-194, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a CSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-100, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting colon Tissue

The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the colon or to specific cells in the colon. In a preferred embodiment, an anti-CSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if colon tissue needs to be selectively destroyed. This would be useful for targeting and killing colon cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting colon cell function.

In another embodiment, an anti-CSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring colon function, identifying colon cancer tumors, and identifying noncancerous colon diseases.

EXAMPLES

Example 1: Gene Expression analysis

Custom Microarray Experiment—Colon Cancer

Custom oligonucleotide microarrays were provided by Agilent Technologies, Inc. (Palo Alto, CA). The microarrays were fabricated by Agilent using their technology for the *in-situ* synthesis of 60mer oligonucleotides. Hughes *et al.*, *Nature Biotechnology* 19:342-347(2001). The 60mer microarray probes were designed by Agilent, from gene sequences provided by diaDexus, using Agilent proprietary algorithms. Whenever possible two differents 60mers were designed for each gene of interest.

All microarray experiments were two-color experiments and were preformed using Agilent-recommended protocols and reagents. Briefly, each microarray was hybridized with cRNAs synthesized from polyA+ RNA, isolated from cancer and normal tissues, labeled with fluorescent dyes Cyanine3 and Cyanine5 (NEN Life Science Products, Inc., Boston, MA) using a linear amplification method (Agilent). In each experiment the experimental sample was polyA+ RNA isolated from cancer tissue from a single individual and the reference sample was a pool of polyA+ RNA isolated from normal tissues of the same organ as the cancerous tissue (*i.e.* normal colon tissue in experiments with colon cancer samples). Hybridizations were carried out at 60°C, overnight using Agilent *in-situ* hybridization buffer. Following washing, arrays were scanned with a

GenePix 4000B Microarray Scanner (Axon Instruments, Inc., Union City, CA). The resulting images were analyzed with GenePix Pro 3.0 Microarray Acquisition and Analysis Software (Axon). Two different chip designs were evaluated with overlapping sets of a total of 38 samples, comparing the expression patterns of colon cancer derived polyA+ RNA to polyA+ RNA isolated from a pool of 7 normal colon tissues. For Chip2 all 38 samples (23 Ascending colon carcinomas and 15 Rectosigmoidal carcinomas including: 5 stage I cancers, 15 stage II cancers, 15 stage III and 2 stage IV cancers, as well as 28 Grade1/2 and 10 Grade 3 cancers) were analyzed. The histopathologic grades for cancer are classified as follows: GX, cannot be assessed; G1, well differentiated; G2, Moderately differentiated; G3, poorly differentiated; and G4, undifferentiated. AJCC Cancer Staging Handbook, 5th Edition, 1998, page 9. For Chip2 analysis, samples were further divided into groups based on the expression pattern of the known colon cancer associated gene Thymidilate Synthase (TS) (13 TS up 25 TS not up). The association of TS with advanced colorectal cancer is well documented. Paradiso *et al.*, *Br J Cancer* 82(3):560-7 (2000); Etienne *et al.*, *J Clin Oncol.* 20(12):2832-43 (2002); Aschele *et al.* *Clin Cancer Res.* 6(12):4797-802 (2000). For Chip 1 a subset of 27 of these samples (14 Ascending colon carcinomas and 13 Rectosigmoidal carcinomas including: 3 stage I cancers, 9 stage II cancers, 13 stage III and 2 stage IV cancers) were assessed.

Data normalization and expression profiling were done with Expressionist software from GeneData Inc. (Daly City, CA/Basel, Switzerland). Gene expression analysis was performed using only experiments that meet certain quality criteria. The quality criteria that experiments must meet are a combination of evaluations performed by the Expressionist software and evaluations performed manually using raw and normalized data. To evaluate raw data quality, detection limits (the mean signal for a replicated negative control + 2 Standard Deviations (SD)) for each channel were calculated. The detection limit is a measure of non-specific hybridization. Acceptable detection limits were defined for each dye (80 for Cy5 and 150 for Cy3). Arrays with poor detection limits in one or both channels were not analyzed and the experiments were repeated. To evaluate normalized data quality, positive control elements included in the array were utilized. These array features should have a mean ratio of 1 (no differential expression). If these features have a mean ratio of greater than 1.5-fold up or down, the experiments were not analyzed further and were repeated. In addition to traditional scatter plots demonstrating the distribution of signal in each experiment, the Expressionist software

also has minimum thresholding criteria that employ user defined parameters to identify quality data. These thresholds include two distinct quality measurements: 1) minimum area percentage, which is a measure of the integrity of each spot and 2) signal to noise ratio, which ensures that the signal being measured is significantly above any background (nonspecific) signal present. Only those features that meet the threshold criteria were included in the filtering and analyses carried out by Expressionist. The thresholding settings employed require a minimum area percentage of 60% [(% pixels > background + 2SD)-(% pixels saturated)], and a minimum signal to noise ratio of 2.0 in both channels. By these criteria, very low expressers, saturated features and spots with abnormally high local background were not included in analysis.

Relative expression data was collected from Expressionist based on filtering and clustering analyses. Up-regulated genes were identified using criteria for the percentage of experiments in which the gene is up-regulated. In general, up-regulation in ~30% of samples tested was used as a cutoff for filtering. The results for Chip 2 are shown in Tables 1, 2, and 3. The results for the statistically significant up-regulated genes on Chip 1 are shown in Table 4.

The first two columns of each table contain information about the sequence itself (Seq ID, Oligo ID), the next columns show the results obtained for all ("ALL") the colon samples, ascending colon carcinomas ("ASC"), Rectosigmoidal carcinomas ("RS"), cancers corresponding to stages I and II ("ST1,2"), stages III and IV ("ST3,4"), grades 1 and 2 ("GR1,2"), grade 3 ("GR3"), cancers exhibiting up-regulation of the TS gene ("TSup") or those not exhibiting up-regulation of the TS gene ("NOT TSup"). '%up' indicates the percentage of all experiments in which up-regulation of at least 2-fold was observed (n=27 for Chip1, n=38 for Chip2), '%valid up' indicates the percentage of experiments with valid expression values in which up-regulation of at least 2-fold was observed.

Table 1. Expression data from Chip2 analysis with Colon cancer samples up-regulated genes.

SeqID	oligoID	%up ALL n=38	%valid up ALL n=38	%up ASC n=23	%valid up ASC n=23	%up RS n=15	%valid up RS n=15
DEX0356_004.nt.1	10720	36.8	36.8	21.7	21.7	60	60
DEX0356_004.nt.1	10721	36.8	36.8	21.7	21.7	60	60
DEX0356_005.nt.1	38049	47.4	47.4	43.5	43.5	53.3	53.3
DEX0356_005.nt.1	38050	55.3	55.3	47.8	47.8	66.7	66.7
DEX0356_006.nt.1	35170	28.9	84.6	21.7	71.4	40	100
DEX0356_007.nt.1	30227	50	50	60.9	60.9	33.3	33.3

DEX0356_007.nt.1	30228	44.7	44.7	56.5	56.5	26.7	26.7
DEX0356_010.nt.1	29571	31.6	31.6	43.5	43.5	13.3	13.3
DEX0356_010.nt.1	29581	50	50	52.2	52.2	46.7	46.7
DEX0356_010.nt.1	29582	52.6	52.6	60.9	60.9	40	40
DEX0356_010.nt.1	29595	52.6	52.6	56.5	56.5	46.7	46.7
DEX0356_010.nt.1	29609	47.4	47.4	52.2	52.2	40	40
DEX0356_010.nt.1	29611	52.6	52.6	56.5	56.5	46.7	46.7
DEX0356_010.nt.1	29612	50	50	52.2	52.2	46.7	46.7
DEX0356_010.nt.2	29581	50	50	52.2	52.2	46.7	46.7
DEX0356_010.nt.2	29582	52.6	52.6	60.9	60.9	40	40
DEX0356_010.nt.2	29595	52.6	52.6	56.5	56.5	46.7	46.7
DEX0356_010.nt.2	29609	47.4	47.4	52.2	52.2	40	40
DEX0356_010.nt.2	29611	52.6	52.6	56.5	56.5	46.7	46.7
DEX0356_010.nt.2	29612	50	50	52.2	52.2	46.7	46.7
DEX0356_010.nt.3	29571	31.6	31.6	43.5	43.5	13.3	13.3
DEX0356_010.nt.3	29581	50	50	52.2	52.2	46.7	46.7
DEX0356_010.nt.3	29582	52.6	52.6	60.9	60.9	40	40
DEX0356_010.nt.3	29595	52.6	52.6	56.5	56.5	46.7	46.7
DEX0356_010.nt.3	29609	47.4	47.4	52.2	52.2	40	40
DEX0356_010.nt.3	29611	52.6	52.6	56.5	56.5	46.7	46.7
DEX0356_010.nt.3	29612	50	50	52.2	52.2	46.7	46.7
DEX0356_013.nt.1	32220	47.4	47.4	43.5	43.5	53.3	53.3
DEX0356_013.nt.1	32221	60.5	63.9	52.2	54.5	73.3	78.6
DEX0356_015.nt.1	33503	18.4	18.9	30.4	31.8	0	0
DEX0356_015.nt.2	33503	18.4	18.9	30.4	31.8	0	0
DEX0356_015.nt.3	33503	18.4	18.9	30.4	31.8	0	0
DEX0356_016.nt.1	35091	21.1	21.1	30.4	30.4	6.7	6.7
DEX0356_016.nt.2	35091	21.1	21.1	30.4	30.4	6.7	6.7
DEX0356_016.nt.3	35091	21.1	21.1	30.4	30.4	6.7	6.7
DEX0356_019.nt.1	35571	18.4	18.4	30.4	30.4	0	0
DEX0356_019.nt.2	35571	18.4	18.4	30.4	30.4	0	0
DEX0356_019.nt.3	35571	18.4	18.4	30.4	30.4	0	0
DEX0356_023.nt.1	8910	26.3	26.3	34.8	34.8	13.3	13.3
DEX0356_023.nt.2	8910	26.3	26.3	34.8	34.8	13.3	13.3
DEX0356_023.nt.3	8910	26.3	26.3	34.8	34.8	13.3	13.3
DEX0356_024.nt.1	36564	34.2	34.2	39.1	39.1	26.7	26.7
DEX0356_025.nt.1	33736	26.3	26.3	34.8	34.8	13.3	13.3
DEX0356_025.nt.1	38517	42.1	42.1	52.2	52.2	26.7	26.7
DEX0356_025.nt.1	38583	47.4	47.4	60.9	60.9	26.7	26.7
DEX0356_025.nt.1	38622	39.5	39.5	52.2	52.2	20	20
DEX0356_025.nt.2	33736	26.3	26.3	34.8	34.8	13.3	13.3
DEX0356_025.nt.2	38517	42.1	42.1	52.2	52.2	26.7	26.7
DEX0356_025.nt.2	38583	47.4	47.4	60.9	60.9	26.7	26.7
DEX0356_025.nt.2	38622	39.5	39.5	52.2	52.2	20	20
DEX0356_026.nt.1	37705	26.3	26.3	21.7	21.7	33.3	33.3
DEX0356_026.nt.1	37706	21.1	21.1	13	13	33.3	33.3
DEX0356_027.nt.1	35470	34.2	34.2	39.1	39.1	26.7	26.7
DEX0356_027.nt.1	35471	42.1	42.1	43.5	43.5	40	40
DEX0356_027.nt.2	35470	34.2	34.2	39.1	39.1	26.7	26.7
DEX0356_027.nt.2	35471	42.1	42.1	43.5	43.5	40	40
DEX0356_029.nt.1	35471	42.1	42.1	43.5	43.5	40	40
DEX0356_031.nt.1	21356	18.4	18.9	30.4	31.8	0	0
DEX0356_031.nt.1	28423	18.4	18.9	30.4	30.4	0	0
DEX0356_031.nt.2	21356	18.4	18.9	30.4	31.8	0	0
DEX0356_032.nt.1	35203	21.1	21.1	30.4	30.4	6.7	6.7
DEX0356_032.nt.1	38805	18.4	18.4	30.4	30.4	0	0
DEX0356_032.nt.1	38806	26.3	26.3	34.8	34.8	13.3	13.3

DEX0356_034.nt.1	30532	44.7	44.7	60.9	60.9	20	20
DEX0356_034.nt.2	30532	44.7	44.7	60.9	60.9	20	20
DEX0356_034.nt.2	34642	23.7	30	30.4	43.8	13.3	14.3
DEX0356_034.nt.2	34643	21.1	24.2	30.4	36.8	6.7	7.1
DEX0356_034.nt.3	34642	23.7	30	30.4	43.8	13.3	14.3
DEX0356_034.nt.3	34643	21.1	24.2	30.4	36.8	6.7	7.1
DEX0356_037.nt.1	37615	47.4	47.4	39.1	39.1	60	60
DEX0356_037.nt.1	37616	44.7	44.7	34.8	34.8	60	60
DEX0356_037.nt.1	37626	36.8	36.8	30.4	30.4	46.7	46.7
DEX0356_037.nt.1	37635	42.1	42.1	34.8	34.8	53.3	53.3
DEX0356_038.nt.1	33741	44.7	47.2	39.1	42.9	53.3	53.3
DEX0356_041.nt.1	35520	39.5	39.5	30.4	30.4	53.3	53.3
DEX0356_041.nt.1	35521	42.1	42.1	34.8	34.8	53.3	53.3
DEX0356_041.nt.1	39581	34.2	36.1	34.8	36.4	33.3	35.7
DEX0356_041.nt.1	39582	42.1	42.1	34.8	34.8	53.3	53.3
DEX0356_043.nt.1	9398	26.3	26.3	30.4	30.4	20	20
DEX0356_043.nt.1	9399	26.3	26.3	30.4	30.4	20	20
DEX0356_043.nt.2	9398	26.3	26.3	30.4	30.4	20	20
DEX0356_043.nt.2	9399	26.3	26.3	30.4	30.4	20	20
DEX0356_045.nt.1	33208	36.8	36.8	52.2	52.2	13.3	13.3
DEX0356_046.nt.1	36669	28.9	28.9	30.4	30.4	26.7	26.7
DEX0356_046.nt.1	36670	28.9	28.9	30.4	30.4	26.7	26.7
DEX0356_047.nt.1	9110	28.9	28.9	30.4	30.4	26.7	26.7
DEX0356_047.nt.1	9111	31.6	31.6	34.8	34.8	26.7	26.7
DEX0356_047.nt.2	9110	28.9	28.9	30.4	30.4	26.7	26.7
DEX0356_047.nt.2	9111	31.6	31.6	34.8	34.8	26.7	26.7
DEX0356_047.nt.3	9110	28.9	28.9	30.4	30.4	26.7	26.7
DEX0356_047.nt.3	9111	31.6	31.6	34.8	34.8	26.7	26.7
DEX0356_047.nt.4	9110	28.9	28.9	30.4	30.4	26.7	26.7
DEX0356_047.nt.4	9111	31.6	31.6	34.8	34.8	26.7	26.7
DEX0356_047.nt.5	9110	28.9	28.9	30.4	30.4	26.7	26.7
DEX0356_047.nt.5	9111	31.6	31.6	34.8	34.8	26.7	26.7
DEX0356_049.nt.1	10297	63.2	63.2	73.9	73.9	46.7	46.7
DEX0356_050.nt.1	30511	23.7	24.3	30.4	31.8	13.3	13.3
DEX0356_053.nt.1	9398	26.3	26.3	30.4	30.4	20	20
DEX0356_053.nt.1	9399	26.3	26.3	30.4	30.4	20	20
DEX0356_054.nt.1	32056	23.7	23.7	34.8	34.8	6.7	6.7
DEX0356_054.nt.2	32056	23.7	23.7	34.8	34.8	6.7	6.7
DEX0356_055.nt.1	37943	52.6	52.6	56.5	56.5	46.7	46.7
DEX0356_055.nt.1	37944	50	51.4	56.5	56.5	40	42.9

Table 2. Expression data from Chip2 analysis with Colon cancer samples up-regulated genes.

SeqID	oligoID	%up ST1,2 n=20	%valid up ST1,2 n=20	%up ST3,4 n=18	%valid up ST3,4 n=18	%up GR1,2 n=28	%valid up GR1,2 n=28	%up GR3 n=10	%valid up GR3 n=10
DEX0356_002.nt.1	31159	5	5	16.7	16.7	0	0	40	40
DEX0356_002.nt.1	34074	5	5.9	22.2	23.5	3.6	4	40	44.4
DEX0356_003.nt.1	36839	10	11.1	33.3	33.3	17.9	18.5	30	33.3
DEX0356_003.nt.2	36839	10	11.1	33.3	33.3	17.9	18.5	30	33.3
DEX0356_003.nt.3	36839	10	11.1	33.3	33.3	17.9	18.5	30	33.3
DEX0356_004.nt.1	10720	30	30	44.4	44.4	42.9	42.9	20	20
DEX0356_004.nt.1	10721	30	30	44.4	44.4	46.4	46.4	10	10
DEX0356_005.nt.1	38049	50	50	44.4	44.4	50	50	40	40
DEX0356_005.nt.1	38050	60	60	50	50	60.7	60.7	40	40
DEX0356_006.nt.1	35170	30	85.7	27.8	83.3	35.7	83.3	10	100

DEX0356_007.nt.1	30227	55	55	44.4	44.4	42.9	42.9	70	70
DEX0356_007.nt.1	30228	45	45	44.4	44.4	35.7	35.7	70	70
DEX0356_008.nt.1	31402	5	5.3	16.7	16.7	3.6	3.7	30	30
DEX0356_009.nt.1	39839	0	0	22.2	22.2	3.6	3.6	30	30
DEX0356_009.nt.1	39840	0	0	22.2	22.2	3.6	3.7	30	30
DEX0356_009.nt.2	39839	0	0	22.2	22.2	3.6	3.6	30	30
DEX0356_009.nt.2	39840	0	0	22.2	22.2	3.6	3.7	30	30
DEX0356_010.nt.1	29571	25	25	38.9	38.9	28.6	28.6	40	40
DEX0356_010.nt.1	29581	45	45	55.6	55.6	50	50	50	50
DEX0356_010.nt.1	29582	45	45	61.1	61.1	46.4	46.4	70	70
DEX0356_010.nt.1	29595	45	45	61.1	61.1	50	50	60	60
DEX0356_010.nt.1	29609	40	40	55.6	55.6	46.4	46.4	50	50
DEX0356_010.nt.1	29611	45	45	61.1	61.1	50	50	60	60
DEX0356_010.nt.1	29612	45	45	55.6	55.6	50	50	50	50
DEX0356_010.nt.2	29581	45	45	55.6	55.6	50	50	50	50
DEX0356_010.nt.2	29582	45	45	61.1	61.1	46.4	46.4	70	70
DEX0356_010.nt.2	29595	45	45	61.1	61.1	50	50	60	60
DEX0356_010.nt.2	29609	40	40	55.6	55.6	46.4	46.4	50	50
DEX0356_010.nt.2	29611	45	45	61.1	61.1	50	50	60	60
DEX0356_010.nt.2	29612	45	45	55.6	55.6	50	50	50	50
DEX0356_010.nt.3	29571	25	25	38.9	38.9	28.6	28.6	40	40
DEX0356_010.nt.3	29581	45	45	55.6	55.6	50	50	50	50
DEX0356_010.nt.3	29582	45	45	61.1	61.1	46.4	46.4	70	70
DEX0356_010.nt.3	29595	45	45	61.1	61.1	50	50	60	60
DEX0356_010.nt.3	29609	40	40	55.6	55.6	46.4	46.4	50	50
DEX0356_010.nt.3	29611	45	45	61.1	61.1	50	50	60	60
DEX0356_010.nt.3	29612	45	45	55.6	55.6	50	50	50	50
DEX0356_011.nt.1	22654	10	10	11.1	11.1	3.6	3.6	30	30
DEX0356_013.nt.1	32220	30	30	66.7	66.7	42.9	42.9	60	60
DEX0356_013.nt.1	32221	50	52.6	72.2	76.5	57.1	59.3	70	77.8
DEX0356_014.nt.1	10992	20	20	11.1	11.1	10.7	10.7	30	30
DEX0356_015.nt.1	33503	15	15.8	22.2	22.2	14.3	14.8	30	30
DEX0356_015.nt.2	33503	15	15.8	22.2	22.2	14.3	14.8	30	30
DEX0356_015.nt.3	33503	15	15.8	22.2	22.2	14.3	14.8	30	30
DEX0356_017.nt.1	39769	5	5	22.2	22.2	7.1	7.1	30	30
DEX0356_017.nt.2	39769	5	5	22.2	22.2	7.1	7.1	30	30
DEX0356_017.nt.3	39769	5	5	22.2	22.2	7.1	7.1	30	30
DEX0356_018.nt.1	31425	5	5.3	16.7	21.4	3.6	4	30	37.5
DEX0356_018.nt.2	31425	5	5.3	16.7	21.4	3.6	4	30	37.5
DEX0356_018.nt.3	31425	5	5.3	16.7	21.4	3.6	4	30	37.5
DEX0356_018.nt.4	31425	5	5.3	16.7	21.4	3.6	4	30	37.5
DEX0356_019.nt.1	35571	25	25	11.1	11.1	14.3	14.3	30	30
DEX0356_019.nt.2	35571	25	25	11.1	11.1	14.3	14.3	30	30
DEX0356_019.nt.3	35571	25	25	11.1	11.1	14.3	14.3	30	30
DEX0356_020.nt.1	32972	0	0	16.7	16.7	0	0	30	30
DEX0356_020.nt.2	32972	0	0	16.7	16.7	0	0	30	30
DEX0356_020.nt.3	32972	0	0	16.7	16.7	0	0	30	30
DEX0356_021.nt.1	10744	15	16.7	16.7	17.6	10.7	11.5	30	33.3
DEX0356_022.nt.1	12057	5	5.3	16.7	17.6	3.6	3.8	30	30
DEX0356_023.nt.1	8910	25	25	27.8	27.8	25	25	30	30
DEX0356_023.nt.2	8910	25	25	27.8	27.8	25	25	30	30
DEX0356_023.nt.3	8910	25	25	27.8	27.8	25	25	30	30
DEX0356_024.nt.1	36564	30	30	38.9	38.9	28.6	28.6	50	50
DEX0356_025.nt.1	33736	20	20	33.3	33.3	21.4	21.4	40	40
DEX0356_025.nt.1	38517	40	40	44.4	44.4	35.7	35.7	60	60
DEX0356_025.nt.1	38583	50	50	44.4	44.4	42.9	42.9	60	60
DEX0356_025.nt.1	38622	35	35	44.4	44.4	32.1	32.1	60	60

DEX0356_025.nt.2	33736	20	20	33.3	33.3	21.4	21.4	40	40
DEX0356_025.nt.2	38517	40	40	44.4	44.4	35.7	35.7	60	60
DEX0356_025.nt.2	38583	50	50	44.4	44.4	42.9	42.9	60	60
DEX0356_025.nt.2	38622	35	35	44.4	44.4	32.1	32.1	60	60
DEX0356_026.nt.1	37705	30	30	22.2	22.2	28.6	28.6	20	20
DEX0356_027.nt.1	35470	30	30	38.9	38.9	32.1	32.1	40	40
DEX0356_027.nt.1	35471	35	35	50	50	42.9	42.9	40	40
DEX0356_027.nt.2	35470	30	30	38.9	38.9	32.1	32.1	40	40
DEX0356_027.nt.2	35471	35	35	50	50	42.9	42.9	40	40
DEX0356_028.nt.1	31545	10	10	16.7	16.7	7.1	7.1	30	30
DEX0356_028.nt.2	31545	10	10	16.7	16.7	7.1	7.1	30	30
DEX0356_029.nt.1	35471	35	35	50	50	42.9	42.9	40	40
DEX0356_031.nt.1	28423	20	20	16.7	17.6	10.7	11.1	40	40
DEX0356_032.nt.1	35203	30	30	11.1	11.1	14.3	14.3	40	40
DEX0356_032.nt.1	38805	30	30	5.6	5.6	14.3	14.3	30	30
DEX0356_032.nt.1	38806	35	35	16.7	16.7	21.4	21.4	40	40
DEX0356_033.nt.1	34888	20	20	16.7	16.7	14.3	14.3	30	30
DEX0356_033.nt.2	34888	20	20	16.7	16.7	14.3	14.3	30	30
DEX0356_033.nt.3	34888	20	20	16.7	16.7	14.3	14.3	30	30
DEX0356_033.nt.4	34888	20	20	16.7	16.7	14.3	14.3	30	30
DEX0356_033.nt.5	34888	20	20	16.7	16.7	14.3	14.3	30	30
DEX0356_033.nt.6	34888	20	20	16.7	16.7	14.3	14.3	30	30
DEX0356_033.nt.7	34888	20	20	16.7	16.7	14.3	14.3	30	30
DEX0356_034.nt.1	30532	50	50	38.9	38.9	35.7	35.7	70	70
DEX0356_034.nt.2	30532	50	50	38.9	38.9	35.7	35.7	70	70
DEX0356_034.nt.2	34642	25	31.2	22.2	28.6	21.4	25	30	50
DEX0356_034.nt.2	34643	20	22.2	22.2	26.7	10.7	12	50	62.5
DEX0356_034.nt.3	34642	25	31.2	22.2	28.6	21.4	25	30	50
DEX0356_034.nt.3	34643	20	22.2	22.2	26.7	10.7	12	50	62.5
DEX0356_035.nt.1	37498	5	5	11.1	11.1	0	0	30	30
DEX0356_036.nt.1	20153	20	21.1	22.2	22.2	14.3	14.8	40	40
DEX0356_037.nt.1	37615	50	50	44.4	44.4	46.4	46.4	50	50
DEX0356_037.nt.1	37616	45	45	44.4	44.4	42.9	42.9	50	50
DEX0356_037.nt.1	37626	30	30	44.4	44.4	32.1	32.1	50	50
DEX0356_037.nt.1	37635	45	45	38.9	38.9	39.3	39.3	50	50
DEX0356_038.nt.1	33741	50	50	38.9	43.8	46.4	48.1	40	44.4
DEX0356_039.nt.1	38976	15	15	22.2	22.2	14.3	14.3	30	30
DEX0356_040.nt.1	38423	15	15	22.2	22.2	14.3	14.3	30	30
DEX0356_040.nt.2	38423	15	15	22.2	22.2	14.3	14.3	30	30
DEX0356_040.nt.3	38423	15	15	22.2	22.2	14.3	14.3	30	30
DEX0356_040.nt.4	38423	15	15	22.2	22.2	14.3	14.3	30	30
DEX0356_040.nt.5	38423	15	15	22.2	22.2	14.3	14.3	30	30
DEX0356_040.nt.6	38423	15	15	22.2	22.2	14.3	14.3	30	30
DEX0356_041.nt.1	35520	40	40	38.9	38.9	46.4	46.4	20	20
DEX0356_041.nt.1	35521	40	40	44.4	44.4	46.4	46.4	30	30
DEX0356_041.nt.1	39581	30	33.3	38.9	38.9	35.7	38.5	30	30
DEX0356_041.nt.1	39582	40	40	44.4	44.4	46.4	46.4	30	30
DEX0356_042.nt.1	33074	10	10	27.8	29.4	7.1	7.4	50	50
DEX0356_042.nt.1	33075	10	10	16.7	16.7	3.6	3.6	40	40
DEX0356_043.nt.1	9398	30	30	22.2	22.2	28.6	28.6	20	20
DEX0356_043.nt.1	9399	30	30	22.2	22.2	28.6	28.6	20	20
DEX0356_043.nt.2	9398	30	30	22.2	22.2	28.6	28.6	20	20
DEX0356_043.nt.2	9399	30	30	22.2	22.2	28.6	28.6	20	20
DEX0356_045.nt.1	33208	40	40	33.3	33.3	32.1	32.1	50	50
DEX0356_046.nt.1	36669	25	25	33.3	33.3	25	25	40	40
DEX0356_046.nt.1	36670	25	25	33.3	33.3	25	25	40	40
DEX0356_047.nt.1	9110	30	30	27.8	27.8	28.6	28.6	30	30

DEX0356_047.nt.1	9111	30	30	33.3	33.3	28.6	28.6	40	40
DEX0356_047.nt.2	9110	30	30	27.8	27.8	28.6	28.6	30	30
DEX0356_047.nt.2	9111	30	30	33.3	33.3	28.6	28.6	40	40
DEX0356_047.nt.3	9110	30	30	27.8	27.8	28.6	28.6	30	30
DEX0356_047.nt.3	9111	30	30	33.3	33.3	28.6	28.6	40	40
DEX0356_047.nt.4	9110	30	30	27.8	27.8	28.6	28.6	30	30
DEX0356_047.nt.4	9111	30	30	33.3	33.3	28.6	28.6	40	40
DEX0356_047.nt.5	9110	30	30	27.8	27.8	28.6	28.6	30	30
DEX0356_047.nt.5	9111	30	30	33.3	33.3	28.6	28.6	40	40
DEX0356_049.nt.1	10297	65	65	61.1	61.1	53.6	53.6	90	90
DEX0356_050.nt.1	30511	20	21.1	27.8	27.8	14.3	14.8	50	50
DEX0356_051.nt.1	41178	10	10	11.1	11.8	3.6	3.7	30	30
DEX0356_052.nt.1	31425	5	5.3	16.7	21.4	3.6	4	30	37.5
DEX0356_053.nt.1	9398	30	30	22.2	22.2	28.6	28.6	20	20
DEX0356_053.nt.1	9399	30	30	22.2	22.2	28.6	28.6	20	20
DEX0356_054.nt.1	32056	25	25	22.2	22.2	14.3	14.3	50	50
DEX0356_054.nt.2	32056	25	25	22.2	22.2	14.3	14.3	50	50
DEX0356_055.nt.1	37943	50	50	55.6	55.6	46.4	46.4	70	70
DEX0356_055.nt.1	37944	50	52.6	50	50	42.9	44.4	70	70

Table 3. Expression data from Chip2 analysis with Colon cancer samples up-regulated genes.

SeqID	oligoID	%up TSup n=13	%valid up TSup n=13	%up NOT TSup n=25	%valid up NOT TSup n=25
DEX0356_002.nt.1	34074	30.8	33.3	4	4.5
DEX0356_003.nt.1	36839	30.8	33.3	16	16.7
DEX0356_003.nt.2	36839	30.8	33.3	16	16.7
DEX0356_003.nt.3	36839	30.8	33.3	16	16.7
DEX0356_004.nt.1	10720	7.7	7.7	52	52
DEX0356_004.nt.1	10721	15.4	15.4	48	48
DEX0356_005.nt.1	38049	61.5	61.5	40	40
DEX0356_005.nt.1	38050	61.5	61.5	52	52
DEX0356_006.nt.1	35170	7.7	33.3	40	100
DEX0356_007.nt.1	30227	61.5	61.5	44	44
DEX0356_007.nt.1	30228	53.8	53.8	40	40
DEX0356_010.nt.1	29571	46.2	46.2	24	24
DEX0356_010.nt.1	29581	69.2	69.2	40	40
DEX0356_010.nt.1	29582	61.5	61.5	48	48
DEX0356_010.nt.1	29595	69.2	69.2	44	44
DEX0356_010.nt.1	29609	61.5	61.5	40	40
DEX0356_010.nt.1	29611	69.2	69.2	44	44
DEX0356_010.nt.1	29612	69.2	69.2	40	40
DEX0356_010.nt.2	29581	69.2	69.2	40	40
DEX0356_010.nt.2	29582	61.5	61.5	48	48
DEX0356_010.nt.2	29595	69.2	69.2	44	44
DEX0356_010.nt.2	29609	61.5	61.5	40	40
DEX0356_010.nt.2	29611	69.2	69.2	44	44
DEX0356_010.nt.2	29612	69.2	69.2	40	40
DEX0356_010.nt.3	29571	46.2	46.2	24	24
DEX0356_010.nt.3	29581	69.2	69.2	40	40
DEX0356_010.nt.3	29582	61.5	61.5	48	48
DEX0356_010.nt.3	29595	69.2	69.2	44	44
DEX0356_010.nt.3	29609	61.5	61.5	40	40
DEX0356_010.nt.3	29611	69.2	69.2	44	44
DEX0356_010.nt.3	29612	69.2	69.2	40	40

DEX0356_012.nt.1	8377	30.8	30.8	4	4
DEX0356_012.nt.2	8377	30.8	30.8	4	4
DEX0356_013.nt.1	32220	53.8	53.8	44	44
DEX0356_013.nt.1	32221	84.6	84.6	48	52.2
DEX0356_014.nt.1	10992	38.5	38.5	4	4
DEX0356_015.nt.1	33503	38.5	38.5	8	8.3
DEX0356_015.nt.2	33503	38.5	38.5	8	8.3
DEX0356_015.nt.3	33503	38.5	38.5	8	8.3
DEX0356_016.nt.1	35091	38.5	38.5	12	12
DEX0356_016.nt.2	35091	38.5	38.5	12	12
DEX0356_016.nt.3	35091	38.5	38.5	12	12
DEX0356_018.nt.1	31425	30.8	30.8	0	0
DEX0356_018.nt.2	31425	30.8	30.8	0	0
DEX0356_018.nt.3	31425	30.8	30.8	0	0
DEX0356_018.nt.4	31425	30.8	30.8	0	0
DEX0356_019.nt.1	33067	30.8	30.8	8	8
DEX0356_019.nt.1	35570	30.8	30.8	8	8.3
DEX0356_019.nt.1	35571	30.8	30.8	12	12
DEX0356_019.nt.2	33067	30.8	30.8	8	8
DEX0356_019.nt.2	35570	30.8	30.8	8	8.3
DEX0356_019.nt.2	35571	30.8	30.8	12	12
DEX0356_019.nt.3	33067	30.8	30.8	8	8
DEX0356_019.nt.3	35570	30.8	30.8	8	8.3
DEX0356_019.nt.3	35571	30.8	30.8	12	12
DEX0356_021.nt.1	10744	30.8	30.8	8	9.1
DEX0356_022.nt.1	12057	30.8	30.8	0	0
DEX0356_023.nt.1	8910	53.8	53.8	12	12
DEX0356_023.nt.2	8910	53.8	53.8	12	12
DEX0356_023.nt.3	8910	53.8	53.8	12	12
DEX0356_024.nt.1	36564	61.5	61.5	20	20
DEX0356_025.nt.1	33736	53.8	53.8	12	12
DEX0356_025.nt.1	38517	69.2	69.2	28	28
DEX0356_025.nt.1	38583	69.2	69.2	36	36
DEX0356_025.nt.1	38622	61.5	61.5	28	28
DEX0356_025.nt.2	33736	53.8	53.8	12	12
DEX0356_025.nt.2	38517	69.2	69.2	28	28
DEX0356_025.nt.2	38583	69.2	69.2	36	36
DEX0356_025.nt.2	38622	61.5	61.5	28	28
DEX0356_026.nt.1	37705	15.4	15.4	32	32
DEX0356_027.nt.1	35470	30.8	30.8	36	36
DEX0356_027.nt.1	35471	38.5	38.5	44	44
DEX0356_027.nt.2	35470	30.8	30.8	36	36
DEX0356_027.nt.2	35471	38.5	38.5	44	44
DEX0356_028.nt.1	31545	30.8	30.8	4	4
DEX0356_028.nt.2	31545	30.8	30.8	4	4
DEX0356_029.nt.1	35471	38.5	38.5	44	44
DEX0356_030.nt.1	8410	30.8	30.8	4	4
DEX0356_030.nt.1	8411	30.8	30.8	4	4
DEX0356_033.nt.1	34888	38.5	38.5	8	8
DEX0356_033.nt.2	34888	38.5	38.5	8	8
DEX0356_033.nt.3	34888	38.5	38.5	8	8
DEX0356_033.nt.4	34888	38.5	38.5	8	8
DEX0356_033.nt.5	34888	38.5	38.5	8	8
DEX0356_033.nt.6	34888	38.5	38.5	8	8
DEX0356_033.nt.7	34888	38.5	38.5	8	8
DEX0356_034.nt.1	30532	61.5	61.5	36	36
DEX0356_034.nt.2	30532	61.5	61.5	36	36

DEX0356_034.nt.2	34642	23.1	33.3	24	28.6
DEX0356_034.nt.2	34643	7.7	10	28	30.4
DEX0356_034.nt.3	34642	23.1	33.3	24	28.6
DEX0356_034.nt.3	34643	7.7	10	28	30.4
DEX0356_036.nt.1	20153	38.5	38.5	12	12.5
DEX0356_037.nt.1	37615	30.8	30.8	56	56
DEX0356_037.nt.1	37616	30.8	30.8	52	52
DEX0356_037.nt.1	37626	30.8	30.8	40	40
DEX0356_037.nt.1	37635	23.1	23.1	52	52
DEX0356_038.nt.1	33741	23.1	27.3	56	56
DEX0356_039.nt.1	38976	15.4	15.4	20	20
DEX0356_040.nt.1	38423	38.5	38.5	8	8
DEX0356_040.nt.2	38423	38.5	38.5	8	8
DEX0356_040.nt.3	38423	38.5	38.5	8	8
DEX0356_040.nt.4	38423	38.5	38.5	8	8
DEX0356_040.nt.5	38423	38.5	38.5	8	8
DEX0356_040.nt.6	38423	38.5	38.5	8	8
DEX0356_041.nt.1	35520	46.2	46.2	36	36
DEX0356_041.nt.1	35521	53.8	53.8	36	36
DEX0356_041.nt.1	39581	46.2	50	28	29.2
DEX0356_041.nt.1	39582	53.8	53.8	36	36
DEX0356_042.nt.1	33074	38.5	38.5	8	8.3
DEX0356_042.nt.1	33075	30.8	30.8	4	4
DEX0356_043.nt.1	9398	30.8	30.8	24	24
DEX0356_043.nt.1	9399	30.8	30.8	24	24
DEX0356_043.nt.2	9398	30.8	30.8	24	24
DEX0356_043.nt.2	9399	30.8	30.8	24	24
DEX0356_044.nt.1	40737	30.8	30.8	4	4
DEX0356_044.nt.2	40737	30.8	30.8	4	4
DEX0356_045.nt.1	33208	61.5	61.5	24	24
DEX0356_046.nt.1	36669	38.5	38.5	24	24
DEX0356_046.nt.1	36670	38.5	38.5	24	24
DEX0356_047.nt.1	9110	23.1	23.1	32	32
DEX0356_047.nt.1	9111	30.8	30.8	32	32
DEX0356_047.nt.2	9110	23.1	23.1	32	32
DEX0356_047.nt.2	9111	30.8	30.8	32	32
DEX0356_047.nt.3	9110	23.1	23.1	32	32
DEX0356_047.nt.3	9111	30.8	30.8	32	32
DEX0356_047.nt.4	9110	23.1	23.1	32	32
DEX0356_047.nt.4	9111	30.8	30.8	32	32
DEX0356_047.nt.5	9110	23.1	23.1	32	32
DEX0356_047.nt.5	9111	30.8	30.8	32	32
DEX0356_048.nt.1	13783	30.8	30.8	8	8
DEX0356_048.nt.1	13784	30.8	30.8	4	4
DEX0356_049.nt.1	10297	84.6	84.6	52	52
DEX0356_050.nt.1	30511	53.8	53.8	8	8.3
DEX0356_051.nt.1	41178	23.1	23.1	4	4.2
DEX0356_052.nt.1	31425	30.8	30.8	0	0
DEX0356_053.nt.1	9398	30.8	30.8	24	24
DEX0356_053.nt.1	9399	30.8	30.8	24	24
DEX0356_054.nt.1	32056	38.5	38.5	16	16
DEX0356_054.nt.2	32056	38.5	38.5	16	16
DEX0356_055.nt.1	37943	76.9	76.9	40	40
DEX0356_055.nt.1	37944	76.9	76.9	36	37.5

Table 4. Expression data from Chip1 analysis with Colon cancer samples up-regulated genes.

SeqID	oligoID	%up ALL n=27	%valid up ALL n=27	%up ASC n=14	%valid up ASC n=14	%up RS n=13	%valid up RS n=13
DEX0356_001.nt.1	2405	14.8	14.8	28.6	28.6	0	0

The location/mapping of the oligos on the colon Cancer genes were as follows:

SeqID	MasterOligoID	oligoID	ChipName	Location
DEX0356_001.nt.1	2405	2405	HGS array	1918-1977
DEX0356_002.nt.1	31159	31159	Colon array	2168-2227
DEX0356_002.nt.1	34074	34074	Colon array	2597-2656
DEX0356_003.nt.1	36839	36839	Colon array	423-478
DEX0356_003.nt.2	36839	36839	Colon array	4021-4076
DEX0356_003.nt.3	36839	36839	Colon array	2523-2578
DEX0356_004.nt.1	10720	10720	Colon array	260-319
DEX0356_004.nt.1	10721	10721	Colon array	220-279
DEX0356_005.nt.1	38049	38049	Colon array	2673-2732
DEX0356_005.nt.1	38050	38050	Colon array	2564-2623
DEX0356_006.nt.1	35170	35170	Colon array	511-570
DEX0356_007.nt.1	30227	30227	Colon array	2563-2622
DEX0356_007.nt.1	30228	30228	Colon array	2517-2576
DEX0356_008.nt.1	31402	31402	Colon array	566-625
DEX0356_009.nt.1	39839	39839	Colon array	975-1034
DEX0356_009.nt.1	39840	39840	Colon array	822-881
DEX0356_009.nt.2	39839	39839	Colon array	1702-1761
DEX0356_009.nt.2	39840	39840	Colon array	1549-1608
DEX0356_010.nt.1	29571	29571	Colon array	79-138
DEX0356_010.nt.1	29581	29581	Colon array	579-634
DEX0356_010.nt.1	29582	29582	Colon array	482-541
DEX0356_010.nt.1	29595	29595	Colon array	493-552
DEX0356_010.nt.1	29609	29609	Colon array	486-545
DEX0356_010.nt.1	29611	29611	Colon array	493-552
DEX0356_010.nt.1	29612	29612	Colon array	456-515
DEX0356_010.nt.2	29581	29581	Colon array	385-440
DEX0356_010.nt.2	29582	29582	Colon array	288-347
DEX0356_010.nt.2	29595	29595	Colon array	299-358
DEX0356_010.nt.2	29609	29609	Colon array	292-351
DEX0356_010.nt.2	29611	29611	Colon array	299-358
DEX0356_010.nt.2	29612	29612	Colon array	262-321
DEX0356_010.nt.3	29571	29571	Colon array	3832-3891
DEX0356_010.nt.3	29581	29581	Colon array	4332-4387
DEX0356_010.nt.3	29582	29582	Colon array	4235-4294
DEX0356_010.nt.3	29595	29595	Colon array	4246-4305
DEX0356_010.nt.3	29609	29609	Colon array	4239-4298
DEX0356_010.nt.3	29611	29611	Colon array	4246-4305
DEX0356_010.nt.3	29612	29612	Colon array	4209-4268
DEX0356_011.nt.1	22654	22654	Colon array	1967-2026
DEX0356_012.nt.1	8377	8377	Colon array	406-465
DEX0356_012.nt.2	8377	8377	Colon array	406-465
DEX0356_013.nt.1	32220	32220	Colon array	1399-1458
DEX0356_013.nt.1	32221	32221	Colon array	1361-1418
DEX0356_014.nt.1	10992	10992	Colon array	122-63
DEX0356_015.nt.1	33503	33503	Colon array	7026-7085
DEX0356_015.nt.2	33503	33503	Colon array	6833-6892
DEX0356_015.nt.3	33503	33503	Colon array	1835-1894
DEX0356_016.nt.1	35091	35091	Colon array	3392-3451
DEX0356_016.nt.2	35091	35091	Colon array	2518-2577

DEX0356_016.nt.3	35091	35091	Colon array	3509-3568
DEX0356_017.nt.1	39769	39769	Colon array	526-585
DEX0356_017.nt.2	39769	39769	Colon array	306-365
DEX0356_017.nt.3	39769	39769	Colon array	393-452
DEX0356_018.nt.1	31425	31425	Colon array	2120-2179
DEX0356_018.nt.2	31425	31425	Colon array	2120-2179
DEX0356_018.nt.3	31425	31425	Colon array	2120-2179
DEX0356_018.nt.4	31425	31425	Colon array	506-565
DEX0356_019.nt.1	33067	33067	Colon array	1441-1500
DEX0356_019.nt.1	35570	35570	Colon array	1437-1495
DEX0356_019.nt.1	35571	35571	Colon array	1385-1444
DEX0356_019.nt.2	33067	33067	Colon array	1441-1500
DEX0356_019.nt.2	35570	35570	Colon array	1437-1495
DEX0356_019.nt.2	35571	35571	Colon array	1385-1444
DEX0356_019.nt.3	33067	33067	Colon array	1441-1500
DEX0356_019.nt.3	35570	35570	Colon array	1437-1495
DEX0356_019.nt.3	35571	35571	Colon array	1385-1444
DEX0356_020.nt.1	32972	32972	Colon array	1059-1118
DEX0356_020.nt.2	32972	32972	Colon array	1718-1777
DEX0356_020.nt.3	32972	32972	Colon array	4421-4480
DEX0356_021.nt.1	10744	10744	Colon array	4464-4523
DEX0356_022.nt.1	12057	12057	Colon array	1271-1330
DEX0356_023.nt.1	8910	8910	Colon array	1898-1957
DEX0356_023.nt.2	8910	8910	Colon array	1022-1081
DEX0356_023.nt.3	8910	8910	Colon array	381-440
DEX0356_024.nt.1	36564	36564	Colon array	315-374
DEX0356_025.nt.1	33736	33736	Colon array	1257-1198
DEX0356_025.nt.1	38517	38517	Colon array	1158-1099
DEX0356_025.nt.1	38583	38583	Colon array	994-935
DEX0356_025.nt.1	38622	38622	Colon array	1196-1137
DEX0356_025.nt.2	33736	33736	Colon array	1257-1198
DEX0356_025.nt.2	38517	38517	Colon array	1158-1099
DEX0356_025.nt.2	38583	38583	Colon array	994-935
DEX0356_025.nt.2	38622	38622	Colon array	1196-1137
DEX0356_026.nt.1	37705	37705	Colon array	273-332
DEX0356_026.nt.1	37706	37706	Colon array	233-292
DEX0356_027.nt.1	35470	35470	Colon array	2787-2846
DEX0356_027.nt.1	35471	35471	Colon array	2663-2721
DEX0356_027.nt.2	35470	35470	Colon array	2833-2892
DEX0356_027.nt.2	35471	35471	Colon array	2709-2767
DEX0356_028.nt.1	31545	31545	Colon array	907-966
DEX0356_028.nt.2	31545	31545	Colon array	938-997
DEX0356_029.nt.1	35471	35471	Colon array	158-100
DEX0356_030.nt.1	8410	8410	Colon array	174-233
DEX0356_030.nt.1	8411	8411	Colon array	144-203
DEX0356_031.nt.1	21356	21356	Colon array	2168-2226
DEX0356_031.nt.1	28423	28423	Colon array	284-343
DEX0356_031.nt.2	21356	21356	Colon array	843-901
DEX0356_032.nt.1	35203	35203	Colon array	134-189
DEX0356_032.nt.1	38805	38805	Colon array	141-200
DEX0356_032.nt.1	38806	38806	Colon array	111-170
DEX0356_033.nt.1	34888	34888	Colon array	965-1024
DEX0356_033.nt.2	34888	34888	Colon array	5326-5385
DEX0356_033.nt.3	34888	34888	Colon array	6110-6169
DEX0356_033.nt.4	34888	34888	Colon array	3084-3143
DEX0356_033.nt.5	34888	34888	Colon array	5785-5844
DEX0356_033.nt.6	34888	34888	Colon array	5605-5664

DEX0356_033.nt.7	34888	34888	Colon array	5301-5360
DEX0356_034.nt.1	30532	30532	Colon array	526-585
DEX0356_034.nt.2	30532	30532	Colon array	525-584
DEX0356_034.nt.2	34642	34642	Colon array	1335-1394
DEX0356_034.nt.2	34643	34643	Colon array	1295-1354
DEX0356_034.nt.3	34642	34642	Colon array	488-547
DEX0356_034.nt.3	34643	34643	Colon array	448-507
DEX0356_035.nt.1	37498	37498	Colon array	409-468
DEX0356_036.nt.1	20153	20153	Colon array	1276-1335
DEX0356_037.nt.1	37615	37615	Colon array	1991-2050
DEX0356_037.nt.1	37616	37616	Colon array	1951-2009
DEX0356_037.nt.1	37626	37626	Colon array	1610-1670
DEX0356_037.nt.1	37635	37635	Colon array	1975-2034
DEX0356_038.nt.1	33741	33741	Colon array	408-467
DEX0356_039.nt.1	38976	38976	Colon array	457-515
DEX0356_040.nt.1	38423	38423	Colon array	232-292
DEX0356_040.nt.2	38423	38423	Colon array	265-325
DEX0356_040.nt.3	38423	38423	Colon array	265-325
DEX0356_040.nt.4	38423	38423	Colon array	467-527
DEX0356_040.nt.5	38423	38423	Colon array	467-527
DEX0356_040.nt.6	38423	38423	Colon array	467-527
DEX0356_041.nt.1	35520	35520	Colon array	655-714
DEX0356_041.nt.1	35521	35521	Colon array	498-557
DEX0356_041.nt.1	39581	39581	Colon array	655-714
DEX0356_041.nt.1	39582	39582	Colon array	545-604
DEX0356_042.nt.1	33074	33074	Colon array	1076-1135
DEX0356_042.nt.1	33075	33075	Colon array	979-1038
DEX0356_043.nt.1	9398	9398	Colon array	303-362
DEX0356_043.nt.1	9399	9399	Colon array	259-318
DEX0356_043.nt.2	9398	9398	Colon array	785-844
DEX0356_043.nt.2	9399	9399	Colon array	741-800
DEX0356_044.nt.1	40737	40737	Colon array	2151-2210
DEX0356_044.nt.2	40737	40737	Colon array	736-795
DEX0356_045.nt.1	33208	33208	Colon array	868-927
DEX0356_046.nt.1	36669	36669	Colon array	6919-6978
DEX0356_046.nt.1	36670	36670	Colon array	6794-6853
DEX0356_047.nt.1	9110	9110	Colon array	206-265
DEX0356_047.nt.1	9111	9111	Colon array	178-231
DEX0356_047.nt.2	9110	9110	Colon array	1624-1683
DEX0356_047.nt.2	9111	9111	Colon array	1596-1649
DEX0356_047.nt.3	9110	9110	Colon array	794-853
DEX0356_047.nt.3	9111	9111	Colon array	766-819
DEX0356_047.nt.4	9110	9110	Colon array	794-853
DEX0356_047.nt.4	9111	9111	Colon array	766-819
DEX0356_047.nt.5	9110	9110	Colon array	794-853
DEX0356_047.nt.5	9111	9111	Colon array	766-819
DEX0356_048.nt.1	13783	13783	Colon array	805-864
DEX0356_048.nt.1	13784	13784	Colon array	765-824
DEX0356_049.nt.1	10297	10297	Colon array	353-412
DEX0356_050.nt.1	30511	30511	Colon array	611-670
DEX0356_051.nt.1	41178	41178	Colon array	519-578
DEX0356_052.nt.1	31425	31425	Colon array	532-591
DEX0356_053.nt.1	9398	9398	Colon array	808-867
DEX0356_053.nt.1	9399	9399	Colon array	764-823
DEX0356_054.nt.1	32056	32056	Colon array	167-108
DEX0356_054.nt.2	32056	32056	Colon array	115-56
DEX0356_055.nt.1	37943	37943	Colon array	522-581

DEX0356_055.nt.1	37944	37944	Colon array	402-461

The mapping of the nucleic acid ("NT") SEQ ID NO; DEX ID; chromosomal location (if known); open reading frame (ORF) location; amino acid ("AA") SEQ ID NO and AA DEX ID are shown in the table below:

NT SEQ_No	NT_SEQID	Cyto_Map	ORF_Loc	AA SEQ_NO	AA_SEQID
1	DEX0356_001.nt.1	6p21.31	296-932	101	DEX0356_001.aa.1
2	DEX0356_002.nt.1	6p22.1	957-1812	102	DEX0356_002.aa.1
3	DEX0356_003.nt.1	19q13.12	1-154	103	DEX0356_003.aa.1
4	DEX0356_003.nt.2	19q13.12	484-1645	104	DEX0356_003.aa.2
5	DEX0356_003.nt.3	19q13.12	100-674	105	DEX0356_003.aa.3
6	DEX0356_004.nt.1	5q32	76-375	106	DEX0356_004.aa.1
7	DEX0356_005.nt.1	8q21.11	1-321	107	DEX0356_005.aa.1
8	DEX0356_006.nt.1	7q31.33	28-330	108	DEX0356_006.aa.1
9	DEX0356_007.nt.1	18q21.33	239-1541	109	DEX0356_007.aa.1
10	DEX0356_008.nt.1	7p15.3	226-922	110	DEX0356_008.aa.1
11	DEX0356_009.nt.1	*	2984-5102	111	DEX0356_009.aa.1
12	DEX0356_009.nt.2	*	549-2262	112	DEX0356_009.aa.2
13	DEX0356_010.nt.1	20p13	311-717	113	DEX0356_010.aa.1
14	DEX0356_010.nt.2	20p13	28-276	114	DEX0356_010.aa.2
15	DEX0356_010.nt.3	20p13	1389-3234	115	DEX0356_010.aa.3
16	DEX0356_011.nt.1	14q23.3	106-1103	116	DEX0356_011.aa.1
17	DEX0356_012.nt.1	17q12	115-646	117	DEX0356_012.aa.1
18	DEX0356_012.nt.2	17q12	115-591	118	DEX0356_012.aa.2
19	DEX0356_013.nt.1	16p13.3	357-1098	119	DEX0356_013.aa.1
20	DEX0356_014.nt.1	6p21.1	133-955	120	DEX0356_014.aa.1
21	DEX0356_015.nt.1	2p23.3	25-6947	121	DEX0356_015.aa.1
22	DEX0356_015.nt.2	2p23.3	25-6755	122	DEX0356_015.aa.2
23	DEX0356_015.nt.3	2p23.3	642-1758	123	DEX0356_015.aa.3
24	DEX0356_016.nt.1	7q11.23	55-2968	124	DEX0356_016.aa.1
25	DEX0356_016.nt.2	7q11.23	204-2094	125	DEX0356_016.aa.2
26	DEX0356_016.nt.3	7q11.23	55-3085	126	DEX0356_016.aa.3
27	DEX0356_017.nt.1	19p13.13	1-576	127	DEX0356_017.aa.1
28	DEX0356_017.nt.2	19p13.13	1-356	128	DEX0356_017.aa.2
29	DEX0356_017.nt.3	19p13.13	1-443	129	DEX0356_017.aa.3
30	DEX0356_018.nt.1	7q22.1	607-2518	130	DEX0356_018.aa.1
31	DEX0356_018.nt.2	7q22.1	607-2655	131	DEX0356_018.aa.2
32	DEX0356_018.nt.3	7q22.1	607-2570	132	DEX0356_018.aa.3
33	DEX0356_018.nt.4	7q22.1	143-1379	133	DEX0356_018.aa.4
34	DEX0356_019.nt.1	6p21.32	1091-2030	134	DEX0356_019.aa.1
35	DEX0356_019.nt.2	6p21.32	1091-1912	135	DEX0356_019.aa.2
36	DEX0356_019.nt.3	6p21.32	1091-1797	136	DEX0356_019.aa.3
37	DEX0356_020.nt.1	10q24.31	270-1053	137	DEX0356_020.aa.1
38	DEX0356_020.nt.2	10q24.31	823-1222	138	DEX0356_020.aa.2
39	DEX0356_020.nt.3	10q24.31	823-1222	138	DEX0356_020.aa.2
40	DEX0356_021.nt.1	13q21.2	229-3811	139	DEX0356_021.aa.1
41	DEX0356_022.nt.1	19q13.42	359-2312	140	DEX0356_022.aa.1
42	DEX0356_023.nt.1	6p24.3	1-495	141	DEX0356_023.aa.1
43	DEX0356_023.nt.2	6p24.3	1-96	142	DEX0356_023.aa.2

44	DEX0356_023.nt.3	6p24.3	1-122	143	DEX0356_023.aa.3
45	DEX0356_024.nt.1	16q22.1	146-569	144	DEX0356_024.aa.1
46	DEX0356_025.nt.1	12p13.33	1-966	145	DEX0356_025.aa.1
47	DEX0356_025.nt.2	12p13.33	1-966	145	DEX0356_025.aa.1
48	DEX0356_026.nt.1	1q21.3	1-222	146	DEX0356_026.aa.1
49	DEX0356_027.nt.1	1p13.3	494-2549	147	DEX0356_027.aa.1
50	DEX0356_027.nt.2	1p13.3	540-2595	147	DEX0356_027.aa.1
51	DEX0356_028.nt.1	2q35	97-1579	148	DEX0356_028.aa.1
52	DEX0356_028.nt.2	2q35	152-1650	149	DEX0356_028.aa.2
53	DEX0356_029.nt.1	1p13.3	46-141	150	DEX0356_029.aa.1
54	DEX0356_030.nt.1	1q32.2	1-265	151	DEX0356_030.aa.1
55	DEX0356_031.nt.1	5q33.1	98-1313	152	DEX0356_031.aa.1
56	DEX0356_031.nt.2	5q33.1	1-309	153	DEX0356_031.aa.2
57	DEX0356_032.nt.1	19q13.42	55-468	154	DEX0356_032.aa.1
58	DEX0356_033.nt.1	19p13.3	58-881	155	DEX0356_033.aa.1
59	DEX0356_033.nt.2	19p13.3	37-1396	156	DEX0356_033.aa.2
60	DEX0356_033.nt.3	19p13.3	37-1858	157	DEX0356_033.aa.3
61	DEX0356_033.nt.4	19p13.3	37-1336	158	DEX0356_033.aa.4
62	DEX0356_033.nt.5	19p13.3	37-1396	156	DEX0356_033.aa.2
63	DEX0356_033.nt.6	19p13.3	37-1396	156	DEX0356_033.aa.2
64	DEX0356_033.nt.7	19p13.3	37-1396	156	DEX0356_033.aa.2
65	DEX0356_034.nt.1	*	3-944	159	DEX0356_034.aa.1
66	DEX0356_034.nt.2	*	73-877	160	DEX0356_034.aa.2
67	DEX0356_034.nt.3	*	163-521	161	DEX0356_034.aa.3
68	DEX0356_035.nt.1	12q13.13	361-676	162	DEX0356_035.aa.1
69	DEX0356_036.nt.1	1q44	76-2383	163	DEX0356_036.aa.1
70	DEX0356_037.nt.1	13q12.3	412-1390	164	DEX0356_037.aa.1
71	DEX0356_038.nt.1	*	88-483	165	DEX0356_038.aa.1
72	DEX0356_039.nt.1	*	560-752	166	DEX0356_039.aa.1
73	DEX0356_040.nt.1	*	367-853	167	DEX0356_040.aa.1
74	DEX0356_040.nt.2	*	163-833	168	DEX0356_040.aa.2
75	DEX0356_040.nt.3	*	1-431	169	DEX0356_040.aa.3
76	DEX0356_040.nt.4	*	602-938	170	DEX0356_040.aa.4
77	DEX0356_040.nt.5	*	602-1233	171	DEX0356_040.aa.5
78	DEX0356_040.nt.6	*	602-1055	172	DEX0356_040.aa.6
79	DEX0356_041.nt.1	*	42-501	173	DEX0356_041.aa.1
80	DEX0356_042.nt.1	19q13.33	1-538	174	DEX0356_042.aa.1
81	DEX0356_043.nt.1	*	109-445	175	DEX0356_043.aa.1
82	DEX0356_043.nt.2	*	244-861	176	DEX0356_043.aa.2
83	DEX0356_044.nt.1	*	55-2283	177	DEX0356_044.aa.1
84	DEX0356_044.nt.2	*	1-867	178	DEX0356_044.aa.2
85	DEX0356_045.nt.1	1q24.2	16-620	179	DEX0356_045.aa.1
86	DEX0356_046.nt.1	2p15	2430-6261	180	DEX0356_046.aa.1
87	DEX0356_047.nt.1	Xq28	82-562	181	DEX0356_047.aa.1
88	DEX0356_047.nt.2	Xq28	1037-1979	182	DEX0356_047.aa.2
89	DEX0356_047.nt.3	Xq28	409-1314	183	DEX0356_047.aa.3
90	DEX0356_047.nt.4	Xq28	1-1179	184	DEX0356_047.aa.4
91	DEX0356_047.nt.5	Xq28	1-873	185	DEX0356_047.aa.5
92	DEX0356_048.nt.1	*	1-299	186	DEX0356_048.aa.1
93	DEX0356_049.nt.1	17q25.3	3-786	187	DEX0356_049.aa.1
94	DEX0356_050.nt.1	19p13.3	88-1375	188	DEX0356_050.aa.1
95	DEX0356_051.nt.1	2p13.2	22-766	189	DEX0356_051.aa.1
96	DEX0356_052.nt.1	7q22.1	139-944	190	DEX0356_052.aa.1
97	DEX0356_053.nt.1	8q22.1	268-970	191	DEX0356_053.aa.1
98	DEX0356_054.nt.1	*	58-287	192	DEX0356_054.aa.1

99	DEX0356_054.nt.2	*	49-235	193	DEX0356_054.aa.2
100	DEX0356_055.nt.1	*14q22.1	1-797	194	DEX0356_055.aa.1

For the polypeptides of the invention, the following attributes were found, epitopes, post translational modifications, signal peptides and transmembrane domains. Antigenicity (Epitope) prediction was performed through the antigenic module in the EMBOSS package. Rice, P., EMBOSS: The European Molecular Biology Open Software Suite, *Trends in Genetics* 16(6): 276-277 (2000). The antigenic module predicts potentially antigenic regions of a protein sequence, using the method of Kolaskar and Tongaonkar. Kolaskar, AS and Tongaonkar, PC., A semi-empirical method for prediction of antigenic determinants on protein antigens, *FEBS Letters* 276: 172-174 (1990). Examples of post-translational modifications (PTMs) and other motifs of the CSPs of this invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. The PTMs and other motifs were predicted by using the ProSite Dictionary of Proteins Sites and Patterns (Bairoch *et al.*, *Nucleic Acids Res.* 25(1):217-221 (1997)), the following motifs, including PTMs, were predicted for the CSPs of the invention. The signal peptides were detected by using the SignalP 2.0, *see* Nielsen *et al.*, *Protein Engineering* 12, 3-9 (1999). Prediction of transmembrane helices in proteins was performed by the application TMHMM 2.0, "currently the best performing transmembrane prediction program", according to authors (Krogh *et al.*, *Journal of Molecular Biology*, 305(3):567-580, (2001); Moller *et al.*, *Bioinformatics*, 17(7):646-653, (2001); Sonnhammer, *et al.*, *A hidden Markov model for predicting transmembrane helices in protein sequences* in Glasgow, *et al.* Ed. Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology, pages 175-182, Menlo Park, CA, 1998. AAAI Press. The PSORT II program may also be used to predict cellular localizations. Horton et al., *Intelligent Systems for Molecular Biology* 5: 147-152 (1997). The table below includes the following sequence annotations: Signal peptide presence; TM (number of membrane domain, topology in orientation and position); Amino acid location and antigenic index (location, AI score, length); PTM and other motifs (type, amino acid residue locations).

SeqID	Signal P	TMHMM	Antigenicity	PTM
DEX0356_001.aa.1	N	0 -o	177-208,1.233; 126-170,1.185;	Bh1 123-141; Bh3 79-93; Ck2_Phospho_Site 26-29, 122-125;

			35-44,1.142; 60-68,1.134; 75-82,1.087; 96-102,1.087; 15-22,1.083; 50-56,1.081; 109-118,1.058	Glycosaminoglycan 5-8; Myristyl 127-132, 180-185; Pkc_Phospho_Site 199-201;
DEX0356_002.aa.1	N	0 -o	127-141,1.201; 155-167,1.188; 255-264,1.165; 81-92,1.153; 40-64,1.145; 18-29,1.144; 96-117,1.136; 215-250,1.119; 188-213,1.111; 177-183,1.094; 143-149,1.09	Asn_Glycosylation 227-230; Camp_Phospho_Site 16-19; Ck2_Phospho_Site 45-48, 76-79, 269-272; Myristyl 14-19, 172-177, 256-261; Pkc_Phospho_Site 21-23, 66-68, 71-73, 229-231;
DEX0356_003.aa.1	N	0 -o	4-18,1.11; 25-37,1.098	Ck2_Phospho_Site 38-41; Myristyl 36-41;
DEX0356_003.aa.2	N	0 -o	206-222,1.204; 103-136,1.177; 230-237,1.162; 184-194,1.16; 145-156,1.141; 245-254,1.139; 367-378,1.132; 14-30,1.129; 331-341,1.115; 308-317,1.104; 160-173,1.091; 95-101,1.075; 196-203,1.072; 79-85,1.049	Asn_Glycosylation 284-287, 353-356; Ck2_Phospho_Site 44-47, 52-55, 176-179, 210-213, 229-232, 230-233, 246-249, 290-293, 306-309, 328-331, 357-360, 358-361; Myristyl 49-54, 57-62, 61-66, 223-228, 275-280, 365-370; Pkc_Phospho_Site 13-15, 23-25, 54-56, 77-79, 90-92, 164-166, 178-180, 210-212, 246-248, 276-278, 374-376, 382-384;
DEX0356_003.aa.3	Y	0 -o	101-121,1.19; 5-46,1.185; 151-164,1.142; 174-187,1.132; 73-86,1.112; 124-145,1.097	Amidation 53-56; Glycosaminoglycan 152-155; Myristyl 67-72, 79-84, 89-94, 95-100, 98-103, 143-148, 153-158;
DEX0356_004.aa.1	Y	0 -o	76-83,1.179; 23-38,1.163;	Amidation 19-22; Camp_Phospho_Site

			5-13,1.148; 58-69,1.138; 89-96,1.075; 49-56,1.063	21-24, 91-94; Ck2_Phospho_Site 46-49; Myristyl 44-49, 73-78; Pkc_Phospho_Site 9-11, 19-21; Kazal 64-86;
DEX0356_005.aa.1	N	0 -o	35-56,1.126; 12-29,1.12; 64-72,1.12; 80-90,1.106	Ck2_Phospho_Site 35-38, 96-99; Myristyl 65-70; Pkc_Phospho_Site 26-28;
DEX0356_006.aa.1	y	1 -o15-37i	27-40,1.171; 7-25,1.14; 48-55,1.118; 74-81,1.098	Asn_Glycosylation 89-92; Camp_Phospho_Site 49-52; Pkc_Phospho_Site 52-54, 92-94;
DEX0356_007.aa.1	N	0 -o	369-384,1.231; 42-54,1.192; 81-98,1.189; 111-162,1.158; 216-227,1.154; 285-293,1.128; 409-417,1.125; 269-278,1.116; 12-29,1.113; 342-355,1.105; 326-335,1.103; 391-403,1.094; 424-430,1.092; 239-245,1.086; 101-108,1.081; 61-67,1.046; 171-177,1.04	Serpin 411-421; Asn_Glycosylation 14-17, 162-165, 196-199, 251-254, 424-427; Ck2_Phospho_Site 82-85, 198-201, 215-218, 304-307; Myristyl 193-198, 270-275, 369-374, 375-380; Pkc_Phospho_Site 35-37, 44-46, 105-107, 148-150, 170-172, 171-173, 198-200;
DEX0356_008.aa.1	N	0 -o	66-79,1.162; 84-103,1.152; 110-117,1.108; 138-148,1.086; 210-217,1.069	Rnp_1 114-121; Amidation 140-143, 184-187; Camp_Phospho_Site 15-18, 20-23, 33-36, 47-50, 60-63, 187-190, 211-214; Ck2_Phospho_Site 85-88, 190-193; Glycosaminoglycan 169-172; Myristyl 160-165, 170-175, 171-176, 172-177,

				173-178, 174-179, 175-180, 176-181, 177-182, 178-183, 179-184, 180-185, 181-186; Pkc_Phospho_Site 7-9, 14-16, 18-20, 31-33, 57-59, 85-87, 110-112, 151-153, 232-234; Tyr_Phospho_Site 186-192, 194-200, 202-208;
DEX0356_009.aa.1	N	1 -o394-416i	551-579,1.243; 37-50,1.233; 341-353,1.232; 391-422,1.231; 18-34,1.218; 158-190,1.213; 54-72,1.198; 81-104,1.19; 207-221,1.189; 306-319,1.186; 667-702,1.182; 622-644,1.18; 455-463,1.17; 607-617,1.17; 518-547,1.166; 276-299,1.146; 425-444,1.136; 378-389,1.136; 247-270,1.13; 192-204,1.118; 355-374,1.115; 117-134,1.11; 500-511,1.095; 485-491,1.077; 233-239,1.076; 588-598,1.073; 139-147,1.061	Asn_Glycosylation 390-393, 643-646; Ck2_Phospho_Site 11-14, 113-116, 307-310, 365-368, 608-611, 645-648; Myristyl 218-223, 249-254, 311-316, 472-477, 473-478, 525-530, 535-540, 570-575, 577-582, 588-593, 699-704; Pkc_Phospho_Site 16-18, 100-102, 112-114, 174-176, 261-263, 294-296, 328-330, 377-379, 487-489, 500-502, 515-517, 583-585, 604-606, 649-651, 683-685; Tyr_Phospho_Site 653-659, 685-692;
DEX0356_009.aa.2	N	10 -i69-86o101-120i240-	71-96,1.26; 544-563,1.229;	Asn_Glycosylation 162-165, 260-263; Ck2_Phospho_Site

		262o267-289i296-318o328-347i359-381o396-418i420-442o452-474i	392-404,1.223; 336-348,1.222; 42-53,1.213; 298-317,1.213; 445-474,1.21; 161-176,1.18; 109-121,1.18; 264-290,1.179; 237-261,1.173; 515-526,1.172; 362-384,1.16; 406-443,1.134; 479-495,1.119; 500-513,1.117; 12-27,1.114; 202-214,1.108; 328-334,1.082; 217-235,1.075; 528-536,1.067; 135-141,1.058; 29-40,1.053; 191-197,1.025	164-167, 186-189, 262-265, 389-392; Myristyl 19-24, 334-339, 409-414, 412-417, 423-428, 542-547; Pkc_Phospho_Site 31-33, 69-71, 180-182, 190-192, 193-195, 324-326, 510-512;
DEX0356_010.aa.1	N	0 -o	82-97,1.201; 30-56,1.184; 68-79,1.102; 119-132,1.08; 100-106,1.044	Amidation 129-132; Asn_Glycosylation 120-123; Camp_Phospho_Site 112-115; Ck2_Phospho_Site 30-33; Myristyl 87-92; Pkc_Phospho_Site 68-70;
DEX0356_010.aa.2	N	0 -o	67-79,1.28; 44-54,1.222; 21-31,1.209; 8-19,1.146; 56-65,1.103	Ck2_Phospho_Site 12-15; Prokar_Lipoprotein 23-33, 61-71;
DEX0356_010.aa.3	N	0 -o	498-524,1.221; 300-312,1.176; 463-479,1.176; 326-335,1.168; 357-370,1.15; 194-	Amidation 66-69; Asn_Glycosylation 58-61; Camp_Phospho_Site 68-71, 389-392; Ck2_Phospho_Site 9-12, 80-83, 81-84, 147-150, 158-161, 187-190, 193-196, 244-247, 319-322,

			219,1.141; 343-355,1.139; 223-240,1.138; 432-442,1.137; 39-53,1.134; 543-558,1.125; 169-183,1.109; 401-414,1.106; 154-162,1.105; 486-492,1.099; 605-611,1.097; 452-460,1.094; 101-115,1.085; 117-124,1.081; 562-574,1.075; 85-93,1.063; 130-140,1.059; 17-25,1.055; 316-322,1.026	394-397, 608-611; Glycosaminoglycan 298-301; Myristyl 74-79, 76-81, 103-108, 107-112, 108-113, 272-277, 275-280; Pkc_Phospho_Site 147-149, 290-292, 464-466, 530-532, 595-597, 608-610; Tyr_Phospho_Site 447-453;
DEX0356_011.aa.1	N	0 -o	49-65,1.181; 25-38,1.149; 108-118,1.147; 236-257,1.142; 122-135,1.132; 193-199,1.131; 267-283,1.128; 160-168,1.125; 297-304,1.121; 81-87,1.12; 68-75,1.118; 175-186,1.11; 5-21,1.1; 311-320,1.098; 215-226,1.064; 99-106,1.048	Amidation 45-48; Camp_Phospho_Site 47-50, 234-237; Ck2_Phospho_Site 83-86, 93-96, 168-171, 181-184, 227-230, 237-240, 245-248, 289-292, 307-310; Myristyl 60-65, 106-111, 207-212, 303-308, 313-318; Pkc_Phospho_Site 121-123, 125-127, 159-161, 231-233, 237-239, 300-302, 316-318;
DEX0356_012.aa.1	N	0 -o	117-142,1.156;	Ck2_Phospho_Site 146-149, 157-160;

			17-23,1.144; 82-98,1.105; 101-112,1.097; 48-76,1.092	Myristyl 14-19, 153-158; Pkc_Phospho_Site 121-123; Proteasome_B 17-64;
DEX0356_012.aa.2	N	0 -o	117-142,1.156; 17-23,1.144; 82-98,1.105; 101-112,1.097; 48-76,1.092	Ck2_Phospho_Site 146-149; Myristyl 14-19; Pkc_Phospho_Site 121-123; Proteasome_B 17-64;
DEX0356_013.aa.1	N	0 -o	65-107,1.225; 27-38,1.2; 113-142,1.174; 201-224,1.153; 47-63,1.144; 4-22,1.133; 153-159,1.123; 174-189,1.118	Trypsin_Ser 170-181; Ck2_Phospho_Site 67-70, 116-119, 195-198, 217-220; Glycosaminoglycan 112-115; Myristyl 46-51, 207-212, 244-249; Pkc_Phospho_Site 52-54, 67-69, 88-90;
DEX0356_014.aa.1	N	0 -o	259-267,1.184; 83-94,1.124; 146-156,1.098; 68-75,1.096; 25-54,1.089; 197-208,1.088; 173-179,1.088; 114-125,1.067; 227-235,1.055	Asn_Glycosylation 24-27, 106-109; Camp_Phospho_Site 274-277; Ck2_Phospho_Site 26-29; Glycosaminoglycan 145-148; Myristyl 16-21, 119-124, 132-137, 146-151, 166-171, 171-176, 214-219, 218-223, 249-254; Pkc_Phospho_Site 20-22, 61-63, 172-174, 245-247, 256-258;
DEX0356_015.aa.1	N	0 -o	1620-1635,1.267; 1074-1095,1.237; 2089-2125,1.236; 79-98,1.233; 1582-1618,1.229; 729-743,1.229; 2165-2191,1.211; 1663-1708,1.207; 1255-1281,1.2; 545-563,1.195; 931-962,1.192; 1419-	Asn_Glycosylation 146-149; Ck2_Phospho_Site 28-31, 108-111, 349-352, 361-364, 374-377, 509-512, 574-577, 644-647, 766-769; 776-779, 803-806, 809-812, 835-838, 929-932, 952-955, 986-989; Myristyl 19-24, 23-28, 85-90, 137-142, 147-152, 189-194, 202-207, 221-226, 255-260, 289-294, 323-328, 327-332, 368-373, 408-413, 476-481, 482-487, 490-495, 566-571, 567-572, 569-574, 615-620, 630-635, 729-734,

			1456,1.19; 1988-2013,1.188; 809-830,1.186; 33-60,1.181; 1134-1155,1.18; 1209-1241,1.18; 1710-1729,1.179; 277-292,1.178; 1458-1475,1.178; 245-265,1.175; 1968-1980,1.173; 874-910,1.171; 1861-1885,1.168; 2295-2303,1.167; 1789-1799,1.167; 609-616,1.164; 688-721,1.164; 176-197,1.161; 446-463,1.158; 1108-1130,1.158; 156-169,1.156; 574-593,1.155; 1181-1198,1.154; 226-242,1.154; 629-638,1.153; 645-672,1.152; 199-212,1.148; 1760-1772,1.147; 1509-1537,1.146; 1565-1575,1.145; 481-505,1.141;	772-777, 900-905; Pkc_Phospho_Site 49-51, 130-132, 168-170, 171-173, 271-273, 290-292, 361-363, 374-376, 380-382, 574-576, 598-600, 757-759, 798-800, 811-813, 877-879, 934-936; Cpsase_1 555-569; Cpsase_2 685-692; Tyr_Phospho_Site 600-607; Gatase_Type_I 252-263;

			1327-1349,1.14; 131-141,1.136; 346-357,1.136; 855-872,1.134; 1354-1365,1.132; 324-340,1.129; 1895-1906,1.129; 387-407,1.129; 843-852,1.129; 2239-2251,1.126; 376-384,1.126; 676-686,1.125; 2195-2203,1.124; 1806-1816,1.121; 1026-1038,1.117; 1828-1836,1.115; 746-765,1.113; 785-797,1.112; 914-921,1.112; 1545-1561,1.111; 2018-2027,1.111; 1389-1395,1.108; 1639-1645,1.106; 146-152,1.106; 1820-1826,1.106; 1296-1303,1.105; 4-15,1.104; 995-1002,1.103; 2069-2076,1.101; 2135-2142,1.098; 17-31,1.094;	

			2270-2277,1.094; 1247-1253,1.094; 601-607,1.093; 2259-2267,1.093; 1157-1171,1.092; 470-476,1.09; 1497-1503,1.089; 1928-1937,1.088; 105-119,1.087; 1313-1320,1.087; 1200-1207,1.085; 1371-1384,1.085; 2219-2226,1.084; 1406-1412,1.084; 1010-1016,1.084; 526-532,1.078; 70-76,1.077; 429-435,1.072; 1051-1058,1.071; 214-220,1.07; 415-423,1.07; 437-444,1.065; 2038-2045,1.055; 1488-1495,1.049; 296-312,1.048; 2079-2085,1.046; 2060-2066,1.046; 1843-1850,1.045; 1941-1947,1.028	
DEX0356_015.aa.2	N	0 -o	1074-1095,1.237; 2025-2061,1.236; 79-98,1.233; 1582-	Asn_Glycosylation 146-149; Ck2_Phospho_Site 28-31, 108-111, 349-352, 361-364, 374-377, 509-512, 574-577,

			1618,1.229; 729-743,1.229; 2101-2127,1.211; 1255-1281,1.2; 545-563,1.195; 931-962,1.192; 1419-1456,1.19; 1924-1949,1.188; 809-830,1.186; 33-60,1.181; 1134-1155,1.18; 1209-1241,1.18; 1629-1648,1.179; 277-292,1.178; 1458-1475,1.178; 245-265,1.175; 1904-1916,1.173; 874-910,1.171; 1780-1804,1.168; 2231-2239,1.167; 1708-1718,1.167; 609-616,1.164; 688-721,1.164; 176-197,1.161; 446-463,1.158; 1108-1130,1.158; 156-169,1.156; 574-593,1.155; 1181-1198,1.154; 226-242,1.154; 629-638,1.153;	644-647, 766-769, 776-779, 803-806, 809-812, 835-838, 929-932, 952-955, 986-989; Myristyl 19-24, 23-28, 85-90, 137-142, 147-152, 189-194, 202-207, 221-226, 255-260, 289-294, 323-328, 327-332, 368-373, 408-413, 476-481, 482-487, 490-495, 566-571, 567-572, 569-574, 615-620, 630-635, 729-734, 772-777, 900-905; Pkc_Phospho_Site 49-51, 130-132, 168-170, 171-173, 271-273, 290-292, 361-363, 374-376, 380-382, 574-576, 598-600, 757-759, 798-800, 811-813, 877-879, 934-936; Cpsase_1 555-569; Cpsase_2 685-692; Tyr_Phospho_Site 600-607; Gatase_Type_I 252-263;

			645-672,1.152; 199-212,1.148; 1679-1691,1.147; 1509-1537,1.146; 1565-1575,1.145; 481-505,1.141; 1327-1349,1.14; 1860-1871,1.139; 131-141,1.136; 346-357,1.136; 855-872,1.134; 1354-1365,1.132; 324-340,1.129; 1814-1825,1.129; 387-407,1.129; 843-852,1.129; 2175-2187,1.126; 376-384,1.126; 676-686,1.125; 2131-2139,1.124; 1620-1627,1.123; 1725-1735,1.121; 1026-1038,1.117; 1747-1755,1.115; 746-765,1.113; 785-797,1.112; 914-921,1.112; 1545-1561,1.111; 1954-1963,1.111; 1389-1395,1.108;	

			146-152,1.106; 1739-1745,1.106; 1296-1303,1.105; 4-15,1.104; 995-1002,1.103; 2005-2012,1.101; 2071-2078,1.098; 17-31,1.094; 2206-2213,1.094; 1247-1253,1.094; 601-607,1.093; 2195-2203,1.093; 1157-1171,1.092; 470-476,1.09; 1497-1503,1.089; 1847-1856,1.088; 105-119,1.087; 1313-1320,1.087; 1200-1207,1.085; 1371-1384,1.085; 2155-2162,1.084; 1406-1412,1.084; 1010-1016,1.084; 526-532,1.078; 70-76,1.077; 429-435,1.072; 1051-1058,1.071; 214-220,1.07; 415-423,1.07; 437-444,1.065; 1974-1981,1.055; 1488-1495,1.049; 296-312,1.048;	

			2015-2021,1.046; 1996-2002,1.046; 1762-1769,1.045	
DEX0356_015.aa.3	N	0 -o	4-11,1.239; 154-190,1.236; 230-256,1.211; 53-78,1.188; 33-45,1.173; 360-368,1.167; 304-316,1.126; 260-268,1.124; 83-92,1.111; 134-141,1.101; 200-207,1.098; 335-342,1.094; 324-332,1.093; 284-291,1.084; 103-110,1.055; 144-150,1.046; 125-131,1.046	Carbamoyltransferase 120-127; Ck2_Phospho_Site 144-147, 157-160, 213-216, 275-278, 282-285, 305-308; Myristyl 34-39, 140-145, 219-224, 223-228; Pkc_Phospho_Site 23-25, 24-26, 97-99, 124-126, 213-215, 251-253, 259-261, 265-267; Tyr_Phospho_Site 356-363;
DEX0356_016.aa.1	N	0 -o	280-297,1.246; 128-151,1.174; 905-940,1.173; 449-555,1.169; 711-725,1.166; 83-104,1.166; 590-600,1.164; 602-619,1.152; 385-399,1.15; 4-20,1.147; 845-863,1.146; 623-671,1.144; 742-757,1.14; 239-255,1.136; 39-61,1.134;	Amidation 692-695; Asn_Glycosylation 768-771; Camp_Phospho_Site 142-145, 195-198, 263-266, 698-701, 734-737, 773-776; Ck2_Phospho_Site 28-31, 30-33, 32-35, 240-243, 266-269, 362-365, 626-629, 630-633, 677-680, 730-733; Glycosaminoglycan 198-201; Leucine_Zipper 826-847; Myristyl 5-10, 12-17, 45-50, 199-204, 700-705, 720-725, 797-802, 862-867, 907-912, 946-951; Pkc_Phospho_Site 38-40, 103-105, 108-110, 502-504, 569-571, 677-679, 732-

			885-898,1.133; 160-167,1.132; 111-119,1.128; 305-320,1.128; 790-810,1.126; 424-446,1.104; 329-339,1.103; 814-822,1.096; 402-409,1.095; 184-191,1.084; 174-180,1.083; 219-234,1.083; 348-358,1.081; 570-581,1.068; 202-210,1.06; 412-421,1.054	734, 750-752, 767-769, 807-809, 916-918, 941-943; Gram_Pos_Anchoring 858-863; Helix_Loop_Helix 811-826;
DEX0356_016.aa.2	N	0 -o	564-599,1.173; 108-214,1.169; 370-384,1.166; 249-259,1.164; 261-278,1.152; 44-58,1.15; 504-522,1.146; 282-330,1.144; 401-416,1.14; 544-557,1.133; 449-469,1.126; 83-105,1.104; 473-481,1.096; 61-68,1.095; 229-240,1.068; 13-20,1.057; 71-80,1.054	Amidation 351-354; Asn_Glycosylation 427-430; Camp_Phospho_Site 357-360, 393-396, 432-435; Ck2_Phospho_Site 285-288, 289-292, 336-339, 389-392; Leucine_Zipper 485-506; Myristyl 17-22, 359-364, 379-384, 456-461, 521-526, 566-571, 605-610; Pkc_Phospho_Site 24-26, 161-163, 228-230, 336-338, 391-393, 409-411, 426-428, 466-468, 575-577, 600-602; Gram_Pos_Anchoring 517-522; Helix_Loop_Helix 470-485;
DEX0356_016.aa.3	N	0 -o	319-336,1.246;	Amidation 731-734; Asn_Glycosylation

			285-298,1.244; 128-151,1.174; 944-979,1.173; 488-594,1.169; 750-764,1.166; 83-104,1.166; 629-639,1.164; 641-658,1.152; 424-438,1.15; 4-20,1.147; 884-902,1.146; 662-710,1.144; 781-796,1.14; 239-263,1.136; 39-61,1.134; 924-937,1.133; 160-167,1.132; 111-119,1.128; 344-359,1.128; 829-849,1.126; 463-485,1.104; 368-378,1.103; 265-271,1.099; 853-861,1.096; 441-448,1.095; 184-191,1.084; 174-180,1.083; 219-234,1.083; 387-397,1.081; 609-620,1.068; 202-210,1.06; 451-460,1.054	807-810; Camp_Phospho_Site 142-145, 195-198, 302-305, 737-740, 773-776, 812-815; Ck2_Phospho_Site 28-31, 30-33, 32-35, 240-243, 275-278, 305-308, 401-404, 665-668, 669-672, 716-719, 769-772; Glycosaminoglycan 198-201; Leucine_Zipper 865-886; Myristyl 5-10, 12-17, 45-50, 199-204, 287-292, 292-297, 739-744, 759-764, 836-841, 901-906, 946-951, 985-990; Pkc_Phospho_Site 38-40, 103-105, 108-110, 541-543, 608-610, 716-718, 771-773, 789-791, 806-808, 846-848, 955-957, 980-982; Rgd 276-278; Gram_Pos_Anchoring 897-902; Helix_Loop_Helix 850-865;
DEX0356_017.aa.1	N	0 -o	83-95,1.18; 4-25,1.16; 100-	Asn_Glycosylation 84-87; Myristyl 40-45, 83-88, 169-174,

			122,1.125; 67-73,1.11; 146-158,1.049	172-177; Pkc_Phospho_Site 86-88; Prokar_Lipoprotein 7-17;
DEX0356_017.aa.2	N	0 -o	49-84,1.09; 104-114,1.087	Myristyl 3-8, 17-22, 27-32, 37-42, 47-52, 55-60, 95-100, 98-103;
DEX0356_017.aa.3	N	0 -o	113-121,1.125; 49-84,1.09	Glycosaminoglycan 112-115; Myristyl 3-8, 17-22, 27-32, 37-42, 47-52, 55-60, 95-100, 98-103, 111-116, 113-118;
DEX0356_018.aa.1	Y	0 -o	295-336,1.235; 8-23,1.223; 597-620,1.201; 36-47,1.199; 240-254,1.178; 338-363,1.159; 559-575,1.156; 226-233,1.15; 444-469,1.135; 108-147,1.127; 384-400,1.124; 412-419,1.107; 501-509,1.105; 192-203,1.102; 170-183,1.099; 577-584,1.097; 259-268,1.096; 478-484,1.077; 523-533,1.076; 512-519,1.074; 52-59,1.073; 376-382,1.07; 491-497,1.068; 63-69,1.064; 151-157,1.056; 624-630,1.038	Amidation 164-167; Asn_Glycosylation 68-71, 553-556; Ck2_Phospho_Site 28-31, 126-129, 372-375, 438-441, 488-491, 490-493, 492-495, 501-504, 524-527; Glycosaminoglycan 7-10; Myristyl 82-87, 172-177, 259-264, 283-288, 593-598, 633-638; Pkc_Phospho_Site 30-32, 70-72, 164-166, 407-409, 476-478, 501-503, 523-525, 638-640; Rgd 473-475; Tyr_Phospho_Site 62-69, 385-393;

DEX0356_018.aa.2	Y	0 -o	295-336,1.235; 8-23,1.223; 36-47,1.199; 240-254,1.178; 338-363,1.159; 559-575,1.156; 649-666,1.151; 226-233,1.15; 444-469,1.135; 108-147,1.127; 384-400,1.124; 412-419,1.107; 501-509,1.105; 192-203,1.102; 170-183,1.099; 577-584,1.097; 259-268,1.096; 612-618,1.086; 478-484,1.077; 523-533,1.076; 512-519,1.074; 638-645,1.074; 52-59,1.073; 376-382,1.07; 491-497,1.068; 63-69,1.064; 151-157,1.056	Amidation 164-167; Asn_Glycosylation 68-71, 553-556, 629-632; Ck2_Phospho_Site 28-31, 126-129, 372-375, 438-441, 488-491, 490-493, 492-495, 501-504, 524-527; Glycosaminoglycan 7-10; Myristyl 82-87, 172-177, 259-264, 283-288, 593-598, 630-635, 637-642; Pkc_Phospho_Site 30-32, 70-72, 164-166, 407-409, 476-478, 501-503, 523-525; Rgd 473-475; Tyr_Phospho_Site 62-69, 385-393;
DEX0356_018.aa.3	Y	0 -o	295-336,1.235; 8-23,1.223; 36-47,1.199; 240-254,1.178; 338-363,1.159; 559-575,1.156; 226-233,1.15; 444-469,1.135;	Amidation 164-167; Asn_Glycosylation 68-71, 553-556; Ck2_Phospho_Site 28-31, 126-129, 372-375, 438-441, 488-491, 490-493, 492-495, 501-504, 524-527, 622-625; Glycosaminoglycan 7-10; Myristyl 82-87, 172-177, 259-264, 283-288, 593-598,

			108-147,1.127; 384-400,1.124; 412-419,1.107; 611-617,1.107; 501-509,1.105; 192-203,1.102; 170-183,1.099; 577-584,1.097; 259-268,1.096; 478-484,1.077; 523-533,1.076; 512-519,1.074; 52-59,1.073; 376-382,1.07; 491-497,1.068; 63-69,1.064; 151-157,1.056	604-609, 617-622, 652-657; Pkc_Phospho_Site 30-32, 70-72, 164-166, 407-409, 476-478, 501-503, 523-525, 653-655; Rgd 473-475; Tyr_Phospho_Site 62-69, 385-393;
DEX0356_018.aa.4	N	0 -o	364-375,1.157; 232-248,1.156; 43-57,1.135; 322-330,1.129; 279-293,1.123; 295-306,1.12; 178-184,1.106; 116-126,1.105; 347-355,1.103; 4-10,1.103; 152-159,1.102; 203-210,1.1; 26-33,1.097; 250-257,1.097; 380-394,1.095; 96-111,1.09; 309-316,1.08; 66-72,1.077; 140-150,1.076;	Amidation 18-21; Asn_Glycosylation 226-229; Camp_Phospho_Site 20-23; Ck2_Phospho_Site 22-25, 23-26, 76-79, 78-81, 80-83, 89-92, 141-144, 180-183, 412-415; Glycosaminoglycan 100-103, 192-195; Myristyl 40-45, 42-47, 46-51, 97-102, 171-176, 204-209, 266-271, 280-285, 399-404; Pkc_Phospho_Site 22-24, 64-66, 89-91, 140-142, 180-182, 192-194, 341-343, 380-382; Rgd 61-63;

			129-136,1.074; 79-92,1.068; 336-343,1.068	
DEX0356_019.aa.1	N	0 -o	261-274,1.217; 225-240,1.194; 37-55,1.146; 83-104,1.142; 58-79,1.128; 282-291,1.119; 145-152,1.111; 110-119,1.102; 206-213,1.1; 194-200,1.074; 249-258,1.043	Amidation 196-199, 264-267; Asn_Glycosylation 216-219, 308-311; Ck2_Phospho_Site 189-192, 205-208; Myristyl 16-21, 52-57, 64-69, 84-89, 133-138, 135-140, 139-144, 159-164, 164-169, 170-175, 172-177, 178-183, 212-217, 248-253; Pkc_Phospho_Site 205-207, 279-281; Tyr_Phospho_Site 74-81; Prokar_Lipoprotein 79-89; Proteasome_B 45-92;
DEX0356_019.aa.2	N	0 -o	37-55,1.146; 83-104,1.142; 58-79,1.128; 145-152,1.111; 110-119,1.102; 206-212,1.1; 253-259,1.097; 194-200,1.074	Amidation 196-199; Camp_Phospho_Site 254-257; Ck2_Phospho_Site 189-192, 205-208; Myristyl 16-21, 52-57, 64-69, 84-89, 133-138, 135-140, 139-144, 159-164, 164-169, 170-175, 172-177, 178-183, 212-217, 259-264; Pkc_Phospho_Site 205-207, 224-226, 253-255; Tyr_Phospho_Site 74-81; Prokar_Lipoprotein 79-89; Proteasome_B 45-92;
DEX0356_019.aa.3	N	0 -o	222-233,1.175; 37-55,1.146; 83-104,1.142; 58-79,1.128; 206-213,1.112; 145-152,1.111; 110-119,1.102; 194-200,1.074	Amidation 196-199; Ck2_Phospho_Site 189-192, 205-208, 228-231; Myristyl 16-21, 52-57, 64-69, 84-89, 133-138, 135-140, 139-144, 159-164, 164-169, 170-175, 172-177, 178-183, 212-217; Pkc_Phospho_Site 205-207; Tyr_Phospho_Site 74-81; Prokar_Lipoprotein 79-89; Proteasome_B

				45-92;
DEX0356_020.aa.1	N	0 -o	238-254,1.196; 53-71,1.181; 20-38,1.144; 40-50,1.103; 97-104,1.099; 128-145,1.093; 223-233,1.092; 163-174,1.09; 5-16,1.088; 198-213,1.076; 151-161,1.056	Ck2_Phospho_Site 21-24, 56-59, 261-264; Leucine_Zipper 53-74, 60-81; Myristyl 90-95, 124-129, 162-167, 165-170; Pkc_Phospho_Site 80-82, 166-168;
DEX0356_020.aa.2	y	0 -o	15-27,1.203; 41-71,1.185; 75-119,1.163; 4-12,1.075; 31-37,1.04	Asn_Glycosylation 127-130; Ck2_Phospho_Site 106-109; Myristyl 36-41, 76-81; Pkc_Phospho_Site 122-124;
DEX0356_021.aa.1	N	0 -o	287-301,1.237; 453-472,1.23; 332-352,1.217; 836-846,1.212; 675-690,1.192; 1097-1106,1.185; 435-450,1.182; 235-253,1.178; 880-893,1.178; 778-810,1.167; 861-871,1.159; 486-500,1.153; 190-214,1.152; 322-330,1.145; 607-619,1.136; 269-280,1.136; 624-635,1.134; 218-229,1.127; 977-999,1.127;	Rnase_Pancreatic 685-691; Asn_Glycosylation 727-730, 868-871; Camp_Phospho_Site 995-998; Ck2_Phospho_Site 53-56, 71-74, 88-91, 123-126, 180-183, 192-195, 199-202, 214-217, 307-310, 318-321, 355-358, 356-359, 432-435, 436-439, 531-534, 830-833, 964-967, 998-1001; Glycosaminoglycan 588-591; Myristyl 19-24, 34-39, 334-339, 557-562, 574-579, 852-857, 858-863; Pkc_Phospho_Site 31-33, 82-84, 88-90, 154-156, 172-174, 203-205, 236-238, 409-411, 432-434, 485-487, 650-652, 709-711, 734-736, 964-966; Tyr_Phospho_Site 416-423, 672-678;

			1180-1190,1.121; 386-394,1.121; 367-380,1.119; 572-600,1.117; 541-565,1.112; 720-735,1.111; 304-313,1.107; 90-106,1.105; 895-917,1.103; 819-830,1.099; 1066-1072,1.099; 755-764,1.088; 1088-1095,1.087; 951-961,1.084; 18-30,1.082; 417-423,1.079; 5-14,1.075; 711-718,1.065; 400-407,1.062; 767-775,1.058; 59-65,1.053; 176-182,1.048	
DEX0356_022.aa.1	N	0 -o	505-537,1.219; 479-503,1.208; 542-554,1.175; 163-172,1.164; 380-396,1.161; 292-308,1.145; 602-614,1.141; 425-449,1.141; 334-348,1.136; 264-286,1.134; 89-102,1.126;	Asn_Glycosylation 103-106, 166-169, 203-206, 223-226, 245-248; Ck2_Phospho_Site 3-6, 27-30, 44-47, 143-146, 185-188, 312-315, 563-566, 622-625; Glycosaminoglycan 584-587; Myristyl 25-30, 181-186, 229-234, 238-243, 241-246, 368-373, 374-379, 403-408, 502-507, 618-623, 629-634; Pkc_Phospho_Site 52-54, 124-126, 143-145, 363-365, 380-382, 424-426, 544-546, 586-588, 622-

			125-140,1.114; 626-637,1.111; 244-254,1.104; 202-208,1.098; 452-473,1.096; 216-226,1.094; 48-54,1.088; 4-12,1.085; 570-576,1.065; 80-87,1.064; 231-237,1.042	624; Tyr_Phospho_Site 564-570;
DEX0356_023.aa.1	y	0 -o	116-135,1.152; 150-159,1.138; 74-89,1.127; 39-48,1.106; 101-114,1.102; 6-23,1.08	Thioredoxin 79-97; Amidation 141-144; Camp_Phospho_Site 143-146; Ck2_Phospho_Site 43-46, 72-75, 96-99, 101-104; Myristyl 20-25, 37-42, 59-64, 110-115; Pkc_Phospho_Site 29-31, 101-103, 129-131;
DEX0356_023.aa.2	N	0 -i	17-28,1.096; 5-13,1.067	Ck2_Phospho_Site 29-32; Pkc_Phospho_Site 16-18; Tyr_Phospho_Site 18-24;
DEX0356_023.aa.3	N	1 -i21-38o	19-36,1.212	
DEX0356_024.aa.1	N	0 -o	111-137,1.283; 4-13,1.181; 94-104,1.167; 35-42,1.129; 65-89,1.119	Amidation 3-6; Ck2_Phospho_Site 57-60; Pkc_Phospho_Site 18-20, 57-59, 62-64;
DEX0356_025.aa.1	y	7 -o20-42i55-77o81-103i120-142o152-171i178-200o224-246i	4-253,1.334; 282-318,1.14; 263-270,1.114	Leucine_Zipper 129-150; Myristyl 12-17, 237-242, 278-283; Pkc_Phospho_Site 15-17, 172-174, 175-177, 299-301; Prokar_Lipoprotein 64-74, 108-118, 110-120, 228-238;
DEX0356_026.aa.1	N	0 -o	19-28,1.168; 62-70,1.085	Thiol_Protease_Asn 40-59; Thiol_Protease_His 23-33; Ck2_Phospho_Site 46-49; Myristyl 33-38;
DEX0356_027.aa.1	N	0 -o	120-132,1.207; 534-	Amidation 408-411, 456-459, 644-647, 682-685;

			542,1.186; 188-195,1.16; 21-28,1.156; 598-607,1.155; 578-585,1.147; 429-450,1.14; 266-274,1.136; 371-378,1.135; 277-289,1.132; 301-313,1.124; 41-56,1.122; 507-514,1.12; 59-93,1.114; 547-554,1.112; 30-39,1.108; 175-184,1.105; 13-19,1.104; 631-638,1.103; 158-165,1.097; 646-652,1.094; 611-619,1.092; 220-233,1.091; 141-156,1.088; 291-299,1.087; 111-117,1.079; 493-499,1.072; 566-573,1.067; 250-256,1.067; 346-352,1.057; 325-335,1.054; 396-402,1.035	Asn_Glycosylation 193-196, 372-375, 387-390, 463-466; Camp_Phospho_Site 410-413, 485-488; Ck2_Phospho_Site 12-15, 28-31, 50-53, 99-102, 265-268, 395-398, 491-494, 612-615, 630-633; Leucine_Zipper 66-87, 73-94; Myristyl 111-116, 211-216, 294-299, 334-339, 381-386, 576-581; Pkc_Phospho_Site 12-14, 116-118, 368-370, 446-448, 465-467, 483-485, 506-508, 561-563, 625-627, 682-684; Tyr_Phospho_Site 243-250;
DEX0356_028.aa.1	N	0 -o	232-253,1.198; 281-316,1.192; 203-212,1.181; 151-165,1.176;	Asn_Glycosylation 92-95; Camp_Phospho_Site 126-129, 344-347; Ck2_Phospho_Site 151-154, 254-257, 319-322, 354-357, 464-467; Myristyl

			328-349,1.163; 466-490,1.147; 169-190,1.143; 48-66,1.139; 401-423,1.136; 114-121,1.132; 255-270,1.118; 90-104,1.116; 364-374,1.115; 4-13,1.111; 126-137,1.111; 436-458,1.108; 382-394,1.102; 29-46,1.098; 273-279,1.084	108-113, 146-151, 147-152, 307-312, 327-332, 473-478; Pkc_Phospho_Site 76-78, 129-131, 170-172, 254-256, 464-466;
DEX0356_028.aa.2	N	0 -o	224-245,1.198; 273-308,1.192; 424-456,1.183; 195-204,1.181; 143-157,1.176; 320-341,1.163; 161-182,1.143; 40-58,1.139; 106-113,1.132; 374-402,1.129; 247-262,1.118; 82-96,1.116; 356-366,1.115; 118-129,1.111; 21-38,1.098; 476-485,1.094; 265-271,1.084; 404-420,1.07; 458-474,1.063	Asn_Glycosylation 84-87, 451-454; Camp_Phospho_Site 118-121, 336-339; Ck2_Phospho_Site 143-146, 246-249, 311-314, 346-349, 428-431, 444-447; Myristyl 100-105, 138-143, 139-144, 299-304, 319-324, 464-469, 470-475; Pkc_Phospho_Site 68-70, 121-123, 162-164, 246-248, 395-397, 425-427;
DEX0356_029.aa.1	N	0 -o	17-28,1.086	Ck2_Phospho_Site 9-12;

DEX0356_030.aa.1	N	0 -i	40-50,1.107; 4-12,1:091; 52-63,1.073; 31-37,1.049	Myristyl 13-18, 19-24, 23-28; Pkc_Phospho_Site 45-47, 69-71, 73-75;
DEX0356_031.aa.1	N	1 -i99-121o	362-401,1.221; 95-114,1.206; 288-305,1.202; 27-38,1.189; 256-277,1.166; 201-211,1.157; 191-199,1.133; 324-332,1.126; 117-125,1.108; 161-167,1.103; 44-50,1.087; 236-243,1.078; 148-155,1.071; 130-138,1.071; 342-348,1.06; 4-12,1.043	Thyroglobulin_1 283-312; Amidation 56-59, 88-91; Asn_Glycosylation 185-188, 191-194, 311-314, 325-328; Camp_Phospho_Site 58-61; Ck2_Phospho_Site 63-66, 80-83, 216-219; Myristyl 93-98, 102-107, 118-123; Pkc_Phospho_Site 63-65, 282-284, 327-329, 335-337, 342-344, 359-361; Tyr_Phospho_Site 286-293;
DEX0356_031.aa.2	N	0 -o	21-31,1.157; 56-63,1.078; 89-95,1.067	Ck2_Phospho_Site 36-39, 98-101; Glycosaminoglycan 93-96; Myristyl 94-99;
DEX0356_032.aa.1	N	0 -o	57-66,1.213; 4-19,1.123; 45-54,1.098; 112-127,1.088; 74-80,1.075	Asn_Glycosylation 17-20; Camp_Phospho_Site 70-73, 117-120, 133-136; Myristyl 55-60; Pkc_Phospho_Site 42-44, 97-99, 125-127;
DEX0356_033.aa.1	N	0 -o	30-57,1.253; 64-102,1.179; 234-266,1.176; 194-213,1.141; 175-192,1.139; 142-149,1.138; 120-136,1.128; 13-20,1.106	Asn_Glycosylation 169-172; Ck2_Phospho_Site 34-37, 48-51, 198-201; Myristyl 79-84, 99-104, 126-131, 206-211; Pkc_Phospho_Site 171-173, 272-274;
DEX0356_033.aa.2	y	0 -o	257-292,1.18; 151-160,1.165; 324-	Asn_Glycosylation 96-99; Camp_Phospho_Site 67-70;

			343,1.153; 35-60,1.152; 121-140,1.141; 102-119,1.139; 69-76,1.138; 189-208,1.13; 224-234,1.126; 300-319,1.121; 236-250,1.11; 5-17,1.091; 167-175,1.08; 371-382,1.074; 384-393,1.073; 411-419,1.055; 403-409,1.05; 26-32,1.049	Ck2_Phospho_Site 25-28, 125-128, 358-361, 447-450; Myristyl 49-54, 65-70, 133-138, 158-163, 166-171, 256-261, 438-443; Pkc_Phospho_Site 14-16, 25-27, 66-68, 98-100, 167-169, 334-336, 434-436; Rgd 443-445; Tyr_Phospho_Site 267-275;
DEX0356_033.aa.3	y	0 -o	257-292,1.18; 448-520,1.174; 395-426,1.154; 324-343,1.153; 35-60,1.152; 434-446,1.148; 550-565,1.144; 121-140,1.141; 102-119,1.139; 166-173,1.139; 69-76,1.138; 189-208,1.13; 151-158,1.129; 224-234,1.126; 536-543,1.124; 300-319,1.121; 576-592,1.115; 236-250,1.11; 5-17,1.091; 373-379,1.05; 26-32,1.049	Asn_Glycosylation 96-99, 487-490; Camp_Phospho_Site 67-70; Ck2_Phospho_Site 25-28, 125-128, 358-361, 466-469; Myristyl 49-54, 65-70, 133-138, 168-173, 256-261, 401-406, 472-477, 528-533, 561-566, 568-573; Pkc_Phospho_Site 14-16, 25-27, 66-68, 98-100, 334-336, 386-388, 394-396, 419-421, 463-465, 506-508, 580-582; Tyr_Phospho_Site 267-275;
DEX0356_033.aa.4	y	0 -o	412-421,1.186; 257-292,1.18;	Asn_Glycosylation 96-99; Camp_Phospho_Site

			151-160,1.165; 324-343,1.153; 35-60,1.152; 121-140,1.141; 373-409,1.139; 102-119,1.139; 69-76,1.138; 189-208,1.13; 224-234,1.126; 300-319,1.121; 236-250,1.11; 5-17,1.091; 167-175,1.08; 26-32,1.049	67-70; Ck2_Phospho_Site 25-28, 125-128, 358-361; Myristyl 49-54, 65-70, 133-138, 158-163, 166-171, 256-261, 412-417, 415-420; Pkc_Phospho_Site 14-16, 25-27, 66-68, 98-100, 167-169, 334-336; Tyr_Phospho_Site 267-275;
DEX0356_034.aa.1	N	0 -o	4-23,1.174; 39-62,1.133; 171-194,1.127; 112-125,1.087; 144-167,1.086; 231-237,1.068; 69-75,1.05	Asn_Glycosylation 222-225; Ck2_Phospho_Site 14-17, 115-118, 122-125, 230-233, 254-257, 279-282, 302-305; Glycosaminoglycan 315-318; Myristyl 67-72, 69-74, 72-77, 85-90, 86-91, 90-95, 91-96, 95-100, 98-103, 104-109, 311-316; Pkc_Phospho_Site 33-35, 82-84, 273-275, 279-281;
DEX0356_034.aa.2	N	0 -o	16-38,1.133; 147-154,1.127; 88-101,1.087; 124-143,1.086; 222-232,1.082; 207-213,1.068; 45-51,1.05; 160-167,1.03	Asn_Glycosylation 198-201, 261-264; Ck2_Phospho_Site 91-94, 98-101, 206-209, 236-239; Myristyl 43-48, 45-50, 48-53, 61-66, 62-67, 66-71, 67-72, 71-76, 74-79, 80-85, 237-242; Pkc_Phospho_Site 7-9, 58-60, 269-271;
DEX0356_034.aa.3	N	0 -o	103-109,1.139; 27-36,1.124; 50-66,1.115; 14-23,1.092	Asn_Glycosylation 116-119; Ck2_Phospho_Site 16-19, 45-48; Myristyl 49-54; Pkc_Phospho_Site 16-18, 72-74, 102-104, 103-105; Tyr_Phospho_Site 18-24, 65-73; If 70-78;
DEX0356_035.aa.1	N	0 -o	64-101,1.148; 5-12,1.07;	Asn_Glycosylation 65-68;

			39-46,1.064	Ck2_Phospho_Site 22-25; Myristyl 54-59; Pkc_Phospho_Site 19-21;
DEX0356_036.aa.1	N	0 -o	544-553,1.231; 370-401,1.222; 227-244,1.195; 27-46,1.178; 152-164,1.163; 284-297,1.141; 572-578,1.138; 269-281,1.132; 10-21,1.128; 348-354,1.125; 585-601,1.125; 439-446,1.119; 534-541,1.114; 419-426,1.104; 496-513,1.103; 692-704,1.1; 359-365,1.093; 759-765,1.083; 321-335,1.08; 407-417,1.075; 67-77,1.072; 521-530,1.068; 299-305,1.062; 125-136,1.043; 48-54,1.032	Atp_Gtp_A 147-154, 452-459; Asn_Glycosylation 531-534, 587-590, 641-644, 745-748; Camp_Phospho_Site 66-69; Ck2_Phospho_Site 95-98, 480-483, 589-592, 601-604; Myristyl 5-10, 27-32, 30-35, 169-174, 173-178, 270-275, 276-281, 317-322, 333-338, 373-378, 452-457, 455-460, 634-639, 652-657, 665-670, 669-674, 670-675, 673-678, 688-693, 689-694, 690-695, 691-696, 692-697, 711-716, 722-727, 738-743, 756-761; Pkc_Phospho_Site 53-55, 81-83, 86-88, 289-291, 330-332, 497-499, 565-567, 621-623, 635-637, 638-640, 643-645, 714-716;
DEX0356_037.aa.1	N	0 -o	4-38,1.178; 55-76,1.156; 262-277,1.145; 105-128,1.122; 242-250,1.118; 165-174,1.106; 188-198,1.097;	Asn_Glycosylation 265-268; Ck2_Phospho_Site 18-21, 142-145, 226-229; Pkc_Phospho_Site 281-283; Tyr_Phospho_Site 107-115, 206-212;

			152-159,1.096	
DEX0356_038.aa.1	N	0 -i	26-41,1.214; 47-70,1.153; 73-84,1.132; 10-20,1.057	Thymosin_B4 109-119; Ck2_Phospho_Site 17-20, 61-64, 92-95; Pkc_Phospho_Site 85-87, 94-96;
DEX0356_039.aa.1	N	0 -o	4-11,1.205; 48-60,1.193; 16-25,1.113; 33-46,1.084	Ck2_Phospho_Site 52-55; Pkc_Phospho_Site 12-14;
DEX0356_040.aa.1	N	0 -o	134-160,1.207; 120-132,1.156; 106-113,1.134; 15-25,1.122	Ck2_Phospho_Site 25-28, 32-35, 106-109; Pkc_Phospho_Site 36-38, 55-57, 121-123, 139-141; Tyr_Phospho_Site 68-74;
DEX0356_040.aa.2	N	0 -o	188-202,1.156; 174-181,1.134; 82-93,1.122; 32-41,1.094; 52-58,1.052; 4-12,1.039	Asn_Glycosylation 221-224; Ck2_Phospho_Site 93-96, 100-103, 174-177; Clathrin_Light_Chn_1 40-46; Myristyl 13-18, 25-30, 27-32, 220-225; Pkc_Phospho_Site 73-75, 104-106, 123-125, 189-191, 207-209; Tyr_Phospho_Site 136-142;
DEX0356_040.aa.3	N	0 -i	36-55,1.186; 16-33,1.186; 86-93,1.094; 127-134,1.06; 106-112,1.052; 58-66,1.039	Amidation 142-145; Camp_Phospho_Site 139-142; Clathrin_Light_Chn_1 94-100; Myristyl 39-44, 41-46, 67-72, 79-84, 81-86; Pkc_Phospho_Site 127-129, 138-140, 142-144;
DEX0356_040.aa.4	N	0 -o	87-105,1.126; 14-25,1.122	Ck2_Phospho_Site 25-28, 32-35; Myristyl 109-114; Pkc_Phospho_Site 36-38, 55-57, 110-112; Tyr_Phospho_Site 68-74, 89-96;
DEX0356_040.aa.5	N	0 -o	146-178,1.243; 201-208,1.162; 120-132,1.156; 106-113,1.134; 15-25,1.122	Amidation 186-189, 189-192, 200-203; Ck2_Phospho_Site 25-28, 32-35, 106-109; Myristyl 208-213; Pkc_Phospho_Site 36-38, 55-57, 121-123, 139-141, 186-188; Tyr_Phospho_Site 68-74;
DEX0356_040.aa.6	N	0 -o	132-147,1.204;	Ck2_Phospho_Site 25-28, 32-35, 106-109;

			106-113,1.134; 15-25,1.122; 119-127,1.056	Pkc_Phospho_Site 36-38, 55-57, 121-123, 122-124, 135-137, 136-138; Tyr_Phospho_Site 68-74;
DEX0356_041.aa.1	Y	0 -o	4-30,1.234; 89-105,1.181; 108-121,1.158; 131-149,1.148; 45-51,1.067; 72-78,1.044	Myristyl 45-50, 95-100; Pkc_Phospho_Site 34-36; Lactalbumin_Lysozyme 100-118;
DEX0356_042.aa.1	N	5 -o29-46i58-75o85-104i117-134o144-166i	15-49,1.224; 141-175,1.216; 58-105,1.201; 113-131,1.153	Amidation 175-178; Camp_Phospho_Site 177-180; Ck2_Phospho_Site 85-88; Myristyl 14-19, 19-24, 59-64, 127-132, 155-160; Pkc_Phospho_Site 63-65; Er_Lumen_Receptor_2 97-106;
DEX0356_043.aa.1	N	1 -o40-62i	31-67,1.264; 74-82,1.212; 4-25,1.199; 85-92,1.154	Asn_Glycosylation 77-80, 88-91; Ck2_Phospho_Site 18-21, 86-89; Pkc_Phospho_Site 57-59; Tyr_Phospho_Site 76-84;
DEX0356_043.aa.2	Y	4 -i26-48o72-94i101-123o155-177i	12-50,1.341; 146-182,1.264; 189-197,1.212; 114-140,1.199; 72-89,1.185; 99-112,1.16	Asn_Glycosylation 59-62, 192-195, 203-206; Ck2_Phospho_Site 61-64, 133-136, 201-204; Pkc_Phospho_Site 172-174; Tyr_Phospho_Site 191-199;
DEX0356_044.aa.1	N	0 -o	238-300,1.255; 6-21,1.176; 615-624,1.174; 729-737,1.119; 384-391,1.114; 47-57,1.112; 493-505,1.11; 104-119,1.109; 331-345,1.109; 187-196,1.107; 144-163,1.095;	Asn_Glycosylation 51-54, 64-67, 434-437, 566-569, 595-598; Camp_Phospho_Site 86-89, 310-313, 667-670, 726-729; Ck2_Phospho_Site 46-49, 54-57, 89-92, 91-94, 96-99, 97-100, 133-136, 134-137, 138-141, 139-142, 173-176, 175-178, 179-182, 180-183, 181-184, 231-234, 232-235, 236-239, 279-282, 313-316, 314-317, 318-321, 319-322, 320-323,

			353-364,1.092; 433-462,1.09; 72-78,1.061; 530-544,1.055; 61-68,1.052; 27-33,1.05; 643-649,1.039; 200-209,1.03; 217-225,1.027	369-372, 373-376, 379-382, 380-383, 381-384, 409-412, 414-417, 418-421, 419-422, 472-475, 473-476, 474-477, 479-482, 480-483, 481-484, 517-520, 519-522, 524-527, 525-528, 526-529, 568-571, 569-572, 573-576, 574-577, 611-614; Myristyl 119-124, 272-277, 275-280, 430-435, 513-518, 548-553, 728-733, 732-737; Pkc_Phospho_Site 66-68, 197-199, 244-246, 253-255, 309-311, 397-399, 431-433, 436-438, 442-444, 466-468, 486-488, 492-494, 498-500, 533-535, 556-558, 586-588, 630-632, 717-719, 723-725, 729-731, 741-743; Prokar_Lipoprotein 270-280;
DEX0356_044.aa.2	N	0 -o	161-170,1.174; 18-48,1.145; 4-10,1.142; 273-283,1.119; 76-90,1.055; 189-195,1.039	Asn_Glycosylation 112-115, 141-144; Camp_Phospho_Site 213-216, 272-275; Ck2_Phospho_Site 18-21, 51-54, 53-56, 63-66, 65-68, 70-73, 71-74, 72-75, 114-117, 115-118, 119-122, 120-123, 157-160; Myristyl 60-65, 94-99, 274-279, 278-283; Pkc_Phospho_Site 79-81, 102-104, 132-134, 176-178, 263-265, 269-271, 275-277, 287-289;
DEX0356_045.aa.1	N	0 -o	103-111,1.202; 134-161,1.182; 69-86,1.159; 43-51,1.124; 191-197,1.094; 94-100,1.09; 121-127,1.074; 20-29,1.068; 170-176,1.063	Atpase_Na_K_Beta_2 48-63; Amidation 116-119; Asn_Glycosylation 60-63, 95-98, 167-170; Ck2_Phospho_Site 188-191; Myristyl 59-64, 135-140; Pkc_Phospho_Site 116-118, 188-190; Rgd 38-40; Tyr_Phospho_Site 179-187;

DEX0356_046.aa.1	N	6 -o342-364i468-490o519-541i553-571o626-648i653-675o	338-369,1.239; 268-301,1.237; 464-506,1.236; 612-659,1.215; 663-675,1.207; 790-818,1.196; 1052-1087,1.186; 306-315,1.184; 415-457,1.183; 73-81,1.18; 188-205,1.179; 30-50,1.175; 217-229,1.175; 978-1000,1.174; 578-585,1.166; 898-932,1.166; 129-148,1.164; 1116-1138,1.163; 1175-1218,1.159; 720-735,1.157; 529-546,1.156; 88-110,1.146; 552-572,1.144; 1021-1037,1.137; 1096-1109,1.131; 825-845,1.13; 112-127,1.127; 150-158,1.126; 743-774,1.125; 235-253,1.121; 1221-1228,1.117; 1145-1158,1.115;	Asn_Glycosylation 136-139, 195-198, 328-331, 380-383, 695-698, 937-940; Camp_Phospho_Site 962-965; Ck2_Phospho_Site 151-154, 220-223, 232-235, 374-377, 378-381, 464-467, 675-678, 842-845, 858-861; Myristyl 29-34, 147-152, 433-438, 696-701, 712-717; Pkc_Phospho_Site 138-140, 291-293, 373-375, 682-684, 687-689, 700-702, 894-896, 957-959, 960-962, 965-967; Prokar_Lipoprotein 528-538;

			400-406,1.115; 1002-1016,1.101; 854-876,1.095; 256-266,1.095; 700-709,1.094; 55-63,1.089; 889-896,1.088; 509-523,1.086; 961-968,1.082; 19-26,1.081; 1258-1265,1.058	
DEX0356_047.aa.1	y	0 -o	4-46,1.193; 137-151,1.11; 98-107,1.055; 83-89,1.037	Asn_Glycosylation 99-102; Camp_Phospho_Site 24-27, 131-134; Ck2_Phospho_Site 60-63; Pkc_Phospho_Site 22-24, 101-103, 115-117;
DEX0356_047.aa.2	N	3 -i72-94o109-131i169-188o	72-96,1.297; 117-137,1.293; 102-115,1.197; 168-200,1.193; 41-51,1.164; 291-305,1.11; 7-14,1.091; 17-25,1.075; 57-63,1.057; 252-261,1.055; 237-243,1.037	Asn_Glycosylation 253-256; Camp_Phospho_Site 28-31, 285-288; Ck2_Phospho_Site 41-44, 214-217; Myristyl 3-8, 38-43, 58-63; Pkc_Phospho_Site 26-28, 70-72, 100-102, 149-151, 255-257, 269-271;
DEX0356_047.aa.3	y	3 -i7-29o44-63i102-121o	5-29,1.297; 50-70,1.293; 277-290,1.265; 224-270,1.239; 35-48,1.197; 101-133,1.193; 185-194,1.055; 170-176,1.037	Asn_Glycosylation 186-189; Camp_Phospho_Site 218-221; Ck2_Phospho_Site 147-150, 283-286; Pkc_Phospho_Site 33-35, 82-84, 188-190, 202-204, 265-267;
DEX0356_047.aa.4	N	2 -o141-163i176-198o	141-165,1.297; 186-206,1.293; 171-	Asn_Glycosylation 322-325; Camp_Phospho_Site 27-30, 354-357; Ck2_Phospho_Site

			184,1.197; 237-269,1.193; 101-119,1.154; 5-14,1.092; 41-47,1.079; 126-132,1.057; 321-330,1.055; 82-89,1.052; 360-379,1.043; 306-312,1.037; 27-35,1.029	283-286; Glycosaminoglycan 74-77; Myristyl 52-57, 70-75, 71-76, 73-78, 75-80, 82-87, 127-132; Pkc_Phospho_Site 79-81, 139-141, 169-171, 218-220, 324-326, 338-340;
DEX0356_047.aa.5	N	3 -o141-163i176-198o239-256i	141-165,1.297; 186-206,1.293; 171-184,1.197; 237-269,1.193; 101-119,1.154; 5-14,1.092; 41-47,1.079; 126-132,1.057; 82-89,1.052; 27-35,1.029	Camp_Phospho_Site 27-30; Ck2_Phospho_Site 283-286; Glycosaminoglycan 74-77; Myristyl 52-57, 70-75, 71-76, 73-78, 75-80, 82-87, 127-132; Pkc_Phospho_Site 79-81, 139-141, 169-171, 218-220;
DEX0356_048.aa.1	N	0 -o	30-38,1.11; 60-84,1.096; 4-12,1.079; 16-22,1.041	Amidation 43-46; Camp_Phospho_Site 54-57; Myristyl 21-26; Pkc_Phospho_Site 25-27, 81-83;
DEX0356_049.aa.1	N	0 -o	105-121,1.209; 23-48,1.204; 60-66,1.121; 219-226,1.11; 179-186,1.105; 133-141,1.104; 149-155,1.064; 247-253,1.054; 75-83,1.044; 123-130,1.041	Asn_Glycosylation 18-21; Ck2_Phospho_Site 117-120, 124-127, 189-192, 247-250; Glycosaminoglycan 99-102; Myristyl 23-28, 26-31, 102-107, 106-111, 110-115, 122-127, 205-210, 237-242; Pkc_Phospho_Site 20-22, 137-139, 148-150;
DEX0356_050.aa.1	N	0 -o	371-383,1.222; 81-99,1.188; 120-155,1.174; 158-176,1.167;	Ck2_Phospho_Site 3-6, 35-38, 114-117, 118-121, 323-326; Myristyl 103-108, 237-242, 241-246; Pkc_Phospho_Site 15-17, 28-30, 61-63,

			300-316,1.166; 324-337,1.16; 399-415,1.154; 227-237,1.151; 212-223,1.117; 351-363,1.107; 246-252,1.107; 178-201,1.106; 389-397,1.093; 268-277,1.076; 23-29,1.068; 203-209,1.065; 284-297,1.059	125-127, 257-259, 287-289, 323-325, 360-362; Lon_Ser 326-334;
DEX0356_051.aa.1	N	0 -o	64-81,1.201; 16-34,1.191; 161-183,1.131; 228-244,1.126; 92-106,1.093; 120-130,1.078; 133-140,1.074; 220-226,1.066; 36-42,1.056	Asn_Glycosylation 138-141; Ck2_Phospho_Site 64-67, 124-127; Myristyl 208-213, 212-217; Pkc_Phospho_Site 121-123;
DEX0356_052.aa.1	N	0 -o	185-202,1.156; 66-96,1.15; 246-266,1.131; 126-137,1.119; 233-240,1.105; 29-35,1.101; 11-27,1.097; 204-211,1.097; 105-111,1.077; 150-160,1.076; 39-46,1.076; 139-146,1.074; 118-124,1.064	Asn_Glycosylation 180-183; Ck2_Phospho_Site 65-68, 115-118, 117-120, 151-154; Myristyl 220-225, 234-239; Pkc_Phospho_Site 150-152; Tyr_Phospho_Site 12-20;
DEX0356_053.aa.1	Y	4 -i26-48o72-	12-50,1.341; 146-	Asn_Glycosylation 59-62, 192-195, 203-

		94i101-123o155-177i	182,1.264; 189-197,1.212; 114-140,1.199; 72-89,1.185; 99-112,1.16; 200-207,1.154; 213-228,1.15	206; Ck2_Phospho_Site 61-64, 133-136, 201-204; Myristyl 219-224; Pkc_Phospho_Site 172-174, 235-237; Tyr_Phospho_Site 191-199;
DEX0356_054.aa.1	N	0 -o	23-61,1.206; 8-18,1.155; 66-73,1.117	Myristyl 9-14, 14-19, 21-26, 30-35;
DEX0356_054.aa.2	N	0 -o	11-46,1.206; 51-58,1.117	Ck2_Phospho_Site 18-21;
DEX0356_055.aa.1	N	0 -o	196-206,1.183; 65-78,1.134; 85-95,1.133; 32-62,1.099; 117-125,1.089; 12-18,1.051; 186-193,1.044	Asn_Glycosylation 133-136; Camp_Phospho_Site 155-158; Ck2_Phospho_Site 9-12, 95-98, 100-103, 149-152; Myristyl 69-74, 256-261; Pkc_Phospho_Site 9-11, 139-141, 149-151, 154-156, 165-167;

Example 2:Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman® probes is a quantitation detection system utilizing the 5'- 3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman®) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently,

first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman® probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the CSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to the calibrator. Normal RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the CSNA in pairs of matched samples may also be determined. A matched pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. All the values are compared to the calibrator.

In the analysis of matching samples, the CSNAs that show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples. Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer state (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matched samples tested are indicative of SEQ ID NO: 1-100 being a diagnostic marker for cancer.

Example 3: Protein Expression

The CSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the CSNA is subcloned in pET-21d for expression in E. coli. In addition to the CSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH_2-terminus of the coding sequence of CSNA, and six histidines, flanking the COOH-terminus of the coding sequence of CSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of CSP is achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that are separated from total cell lysate were incubated with a nickel chelating resin. The column is packed and washed with five column volumes of wash buffer. CSP is eluted stepwise with various concentration imidazole buffers.

Example 4: Fusion Proteins

The human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5'and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e.g.*, WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention;

however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1-100. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al.*, *Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons are also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al.*, *Nucleic Acids Res.*, 19: 1156

(1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Manheim), and FISH is performed as described in Johnson *et al.*, *Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. Johnson (1991). Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 :l of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 :l of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1, μg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μg/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), bucally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481, the contents of which are hereby incorporated by reference herein in their entirety), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324, the contents of which are hereby incorporated by reference herein in their entirety. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably, the carrier is a parenteral carrier, more preferably, a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense or RNAi technology are used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of

the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5'and 3'end sequences respectively as set forth in Example 3. Preferably, the 5'primer contains an EcoRI site and the 3'primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector

that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 12: Method of Treatment Using Gene Therapy-In Vivo

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, Tabata H. *et al. Cardiovasc. Res.* 35 (3): 470-479 (1997); Chao J *et al.* Pharmacol. Res. 35 (6): 517-522 (1997); Wolff J. A. *Neuromuscul. Disord.* 7 (5): 314-318 (1997), Schwartz B. *et al. Gene Ther.* 3 (5): 405-411 (1996); and Tsurumi Y. *et al. Circulation* 94 (12): 3281-3290 (1996); W0 90/11092, W0 98/11779; U. S. Patent No. 5,693,622; 5,705,151; 5,580,859, the contents of which are hereby incorporated by reference herein in their entirety.

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, colon, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. *et al. Ann. NY Acad. Sci.* 772: 126-139 (1995) and Abdallah B. *et al. Biol. Cell* 85 (1): 1-7 (1995)) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art

can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, colon, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. In vivo muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 μg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to colons or bronchial tissues, throat or mucous membranes of the nose. In addition, naked

polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle in vivo is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e.g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (I. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver et al., *Biotechnology* 11: 1263-1270 (1993); Wright et al., *Biotechnology* 9: 830-834 (1991); and U. S. Pat. No. 4,873,191, the contents of which is hereby incorporated by reference herein in its entirety); retrovirus mediated gene transfer into germ lines (Van der Putten et al., *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, *Mol Cell. Biol.* 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., *Science* 259: 1745 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., *Cell* 57: 717-723 (1989). For a review of such techniques, see Gordon, "Transgenic Animals," *Intl. Rev. Cytol.* 115: 171-229 (1989).

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell *et al., Nature* 380: 64-66 (1996); Wilmut *et al., Nature* 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e.g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene

may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., *Science* 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503512 (1987); Thompson et al., Cell 5: 313-321 (1989)) Alternatively, RNAi technology may be used. For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention in vivo. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, supra). However, this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site in vivo using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient in vivo. Such cells may be obtained from the patient (i.e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered in vitro using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e.g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent No. 5,399,349; and Mulligan & Wilson, U. S. Patent No. 5,460,959, the contents of which are hereby incorporated by reference herein in their entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

SEQUENCES

>DEX0356_001.nt.1

CATCCGGGTGGCCACAGAGCAACTTCCTCTAGAGGGAGCTGATTGGAGCCGGGTGCCGCTGGCACCTCTA
TGATCACTGGAGTCTCGCGGGTCCCTCGGGCTGCACAGGGACAAGTAAAGGCTACATCCAGATGCCGGGA
ATGCACTGACGCCCATTCCTGGAAACTGGGCTCCCACTCAGCCCCTGGGAGCAGCAGCCGCCAGCCCCTC
GGGACCTCCATCTCCACCCTGCTGAGCCACCCGGGTTGGGCCAGGATCCCGGCAGGCTGATCCCGTCCTC
CACTGAGACCTGAAAAATGGCTTCGGGGCAAGGCCCAGGTCCTCCCAGGCAGGAGTGCGGAGAGCCTGCC
CTGCCCTCTGCTTCTGAGGAGCAGGTAGCCCAGGACACAGAGGAGGTTTTCCGCAGCTACGTTTTTTACC
GCCATCAGCAGGAACAGGAGGCTGAAGGGGTGGCTGCCCCTGCCGACCCAGAGATGGTCACCTTACCTCT
GCAACCTAGCAGCACCATGGGGCAGGTGGGACGGCAGCTCGCCATCATCGGGGACGACATCAACCGACGC
TATGACTCAGAGTTCCAGACCATGTTGCAGCACCTGCAGCCCACGGCAGAGAATGCCTATGAGTACTTCA
CCAAGATTGCCACCAGCCTGTTTGAGAGTGGCATCAATTGGGGCCGTGTGGTGGCTCTTCTGGGCTTCGG
CTACCGTCTGGCCCTACACGTCTACCAGCATGGCCTGACTGGCTTCCTAGGCCAGGTGACCCGCTTCGTG
GTCGACTTCATGCTGCATCACTGCATTGCCCGGTGGATTGCACAGAGGGGTGGCTGGGTGAGTATCCAAG
GACTGCAATGTCCTCCCTGCTGTTGGGGCTGCCCCTCTCCCAGGCCCCACCCCTTCCTCGGGTTCCTATA
TCTTCTGCTAGCCTCTCTGTGACCACCTTGTTTCTCCCGGCAGGTGGCAGCCCTGAACTTGGGCAATGGT
CCCATCCTGAACGTGCTGGTGGTTCTGGGTGTGGTTCTGTTGGGCCAGTTTGTGGTACGAAGATTCTTCA
AATCATGACTCCCAAGGGTGCCCTTTGGGGTCCCGGTTCAGACCCCTGCCTGGACTTAAGCGAAGTCTTT
GCCTTCTCTGCTCCCTTGCAGGGGTCCCCCCTCAAGAGTACAGAAGCTTTAGCAAGTGTGCACTCCAGCT
TCGGAGGGCCCCTGCGTGGGGGCCAGTCAGGCTGCAGAGGCACCTCAACATTGCATGGTGCTAGTGGGCC
CTCTCTCTGGGCCCAGGGGCTGTGGCCGTCTCCTCCCTCAGCTCTCTGGGACCTCCTTAGCCCTGTCTGC
TAGGCGCTGGGGAGACTGATAACTTGGGGAGGCAAGAGACTGGGAGCCACTTCTCCCCAGAAAGTGTTTA
ACGGTTTTAGCTTTTTATAATACCCTTGTGAGAGCCCATTCCCACCATTCTACCTGAGGCCAGGACGTCT
GGGGTGTGGGGATTGGTGGGTCTATGTTCCCCAGGATTCAGCTATTCTGGAAGATCAGCACCCTAAGAGA
TGGGACTAGGACCTGAGCCTGGTCCTGGCCGTCCCTAAGCATGTGTCCCAGGAGCAGGACCTACTAGGAG
AGGGGGGCCAAGGTCCTGCTCAACTCTACCCCTGCTCCCATTCCTCCCTCCGGCCATACTGCCTTTGCAG
TTGGACTCTCAGGGATTCTGGGCTTGGGGTGTGGGGTGGGGTGGAGTCGCAGACCAGAGCTGTCTGAACT
CACGTGTCAGAAGCCTCCAAGCCTGCCTCCCAAGGTCCTCTCAGTTCTCTCCCTTCCTCTCTCCTTATAG
ACACTTGCTCCCAACCCATTCACTACAGGTGAAGGCTCTCACCCCCATCCCTGGGGGCCTTGGGTGAGTG
GCCTGCTAAGGCTCCTCCTTGCCCAGACTACAGGGCTTAGGACTTGGTTTGTTATTTCAGGGAAAAGGAG
TAGGGAGTTCATCTGGAGGGTTCTAAGTGGGAGAAGGACTATCAACACCACTAGGAATCCCAGAGGTGGG
ATCCTCCCTCATGGCTCTGGCACAGTGTAATCCAGGGGTGTAGATGGGGGAACTGTGAATACTTGAACTC
TGTTCCCCCACCCTCCATGCTCCTCACCTGTCTAGGTCTCCTCAGGGTGGGGGGTGACAGTGCCTTCTCT
ATTGGGCACAGCCTAGGGTCTTGGGGGTCAGGGGGGAGAAGTTCTTGATTCAGCCAAATGCAGGGAGGGG
AGGCAGATGGAGCCCATAGGCCACCCCCTATCCTCTGAGTGTTTGGAAATAAACTGTGCAATCCCCTCA

>DEX0356_002.nt.1

TTTTTTTCATTTTTCTTTTTTGTAGAGGCAGGGTCTTACCATGTTGCCCAGGCTGGTCTTGAACTCCTAG
CCTCAAAGTAATTATCCCACCTTGGCCTCCAAAAGTGTTGGGATTGCAGGCATGAGCCACTGTTTCCGGT
CATGTTACATTTTTAAGTAGATTTCTCAAAACACCATGAGATACGTCTTACTTCTCATAGGCCAGAACTT
AGTCACAAGGTCACATCTAGCTATGAGGGACTGGGCAGCTGTGAGCCATAAAAGGCAGAAGATGGCAGAA
GGGGTATTGGATAGGCAACTGGCAGTCTTTGCTATAGACATCCTTGCAGGGCTAGGGAATTGGGTAATGT
CCCCCACAGTGCCTTTCTATCTGAACATGTCATTCTTTCCATATTTGGGAACCTGAGAATTCGGGTTAAT
TGCCCACCATAAATATCAAGGACTGTCATGGAATCTAGTGCTTTCTGGCAAAGTCAGCCTCATATGTAAT
TCAGTAAACCTTTTTGAGTACTTAGTATTTGAAAAGCTGTGAGCTAGGTGCTTTGGGGGGTATGAGAATG
AGTACAGTATCCTATTTTGGTCTAGAGGATCATAGCTCAGGAAGAGAAGGTGGGAAAAGCTGTAGTGGCT
GATTTTAAAGACTGTCCGTGTACGCTAGACAGTTTAACTTAGCCCCTGTCCTAGTTCCTGGAAGGTCATG
GAATTGGGAAAAAGAATTTACTTGGGCCGGACGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGG
CCGAGGCGGGAGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGACCAGTATGGAGAAACCCCATCTC
TACTAAAAACACAAAATTAGCTGGGTATGGTGGCGCATGCCTGTAATCCCAGCTACTCAGGAGGCTGAGG
CAGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTGCGGTGGGCCGAGATGGCGCCATTGCACTCCAGCCT
GGGCAACAAGCGAAACTCCATCTCAAAAAAAAAAAAAAATTTACTCATGTTTTTTAAATTTGGCTGAAAGA
ATCAGGATTCCTGAACCTTGGATCACACCTCCAGATTTGCAAGAGAAAATCCACATTTTTGCCCAAAAAT
GTCTATTCTTGACGGAGAGTCTAAAGCAGTTCACAGAAAAAATGCAGTCAGATATGGAGAAAATCCAAGA
ATTAAGAGAGGCTCAGTTATACTCAGTGGACGTGACTCTGGACCCAGACACGGCCTACCCCAGCCTGATC
CTCTCTGATAATCTGCGGCAAGTGCGGTACAGTTACCTCCAACAGGACCTGCCTGACAACCCCGAGAGGT
TCAATCTGTTTCCCTGTGTCTTGGGCTCTCCATGCTTCATCGCCGGGAGACATTATTGGGAGGTAGAGGT
GGGAGATAAAGCCAAGTGGACCATAGGTGTCTGTGAAGACTCAGTGTGCAGAAAAGGTGGAGTAACCTCA
GCCCCCCAGAATGGATTCTGGGCAGTGTCTTTGTGGTATGGGAAAGAATATTGGGCTCTTACCTCCCCAA

TGACTGCCCTACCCCTGCGGACCCCGCTCCAGCGGGTGGGGATTTTCTTGGACTATGATGCTGGTGAGGT
CTCCTTCTACAACGTGACAGAGAGGTGTCACACCTTCACTTTCTCTCATGCTACCTTTTGTGGGCCTGTC
CGGCCCTACTTCAGTCTGAGTTACTCGGGAGGGAAAAGTGCAGCTCCTCTGATCATCTGCCCCATGAGTG
GGATAGATGGGTTTTCTGGCCATGTTGGGAATCATGGTCATTCCATGGAGACCTCCCCTTGAGGAGGTGA
ATTCAGGCCAAAAGGGCTGTTGGCTGTAATCCTACGCCAGGCACAAGGCATCTTGTTGCCTTGCCACGTC
CTGTCACAGCTGGGTATCCTTACCATGTTCCACGCCCTTGCAGTGGGAGACAGGATGTCCATGTTCTCTA
CCATCCTTTTCCTTCCCATGCAGATTGTGAAATGTAATGAGATGTATCAAGATATCCTAGAAATAAAAAC
CAGATGTCCACCTCCAGTGTTTCATACTTTCTGGTTTTACACATCGCTGGAGGGATAAAGAGTATGGATA
ATCTTTGGATTTGGAGAGCCGTTCAAGATACTTCCAGCTTCTTGGCTCAGCCTGGCTTCCTCTGGTTCAG
CCCCACATAATGATTATGGCTATTTGCTGTCATTTCTGGGCTAGGGCTCCTTTCTAACAACCTAGACTGG
AATAAGGCCCTGTCAGCATGGCTCCCTTTATCCCAGTTTTCCGTCTGGGAACAGTACCTCTGCCCCTGAT
TCCCAATGTGCCATAGTTTTATTAACTCCATTAAAGAAGCCTGTATGTGTTTTGGTTAGTTACAGTTATT
TTACAATAATGGTGGGTAATGGCCCCACCTCTGTTATGAGATAATGTTCTAATCAATGTCTCTGCCTTTG
TATCTTTTCTGAGGGCTTTGTCTGTTCTCTTCATTCTAATGAAAGGTGTATTCTAGTGCTGGGTGCATAT
CATCCAGGATAATATTCTGCCCAACTCCATCCTCTGTTACTAGATCCCTTACCAGTCACATTTGTGGACT
GGTGGCCAGTCGTATACCATCCCTGGAAGGATTCTGGGACAATATTCCAGGGATTCATTGACTTCTTGGC
TCCTTTTCTCCATTTCCTTTGGGGGAAGGGGGAATTGACCATGCTTAAGTGCATCCTATCAAGGGGCAGC
TCCGTCCCCATGGCCATTGGATCATGAGACACTCTGAAGTCAGAAGGCTGGGGCAGATCACTTCAAGCAA
GCCCCCATGATGGTTCTCAGTCCTGCTTCTCTGTGGGTACGTGCCCCTCTGTTTAAAAATAAACTGAATA
TGGATGTTTA

>DEX0356_003.nt.1
TACTATATACGGTCTAAGGTAGCGAAGATCTGCGGGCTCGTGGAGCTCCTGGATGAGATGAAGTTCTCGC
TGGAGAAGCTGCACCAAGGCATCACAGTCTCAGACCCTCCCTTTGACACCCAGCCCCGGCCCGATGACAG
CTTTTCCTGAGGACCCCGGCCACGCAGCTGTTCCCCCACATGGACAGATGGACACACAGAGCCTCGGCGG
CCACTGCTGGCACGGTGTGAGCGCCAGGCATCTCCCACCCGCCCCTCCCGACGGCCCAACCAGGGGCTGT
GCAGACGTGGGGACCACGGAACCGAGATGCACTTTAGACCAGGGAGCTGGCCCGGCCTCTGGCAGGCCCC
CCACTAACTTATTTTGCCCGGCTGAGGTTGTGGGGGGCGCCTCCTGGGGTGCACGATTCCCTCAGCTCTG
GGTTTAATGTATTATATTTATTTGGGGCCGACAGTGCCCCAATAAAGGGTCAGAAGTG

>DEX0356_003.nt.2
GTCCAGGAGCTTCTGGCCAGGGCATAGTAGTTGCTAGGGACCCTTTTCTAGGTACTAGGGGAGGGGGCTC
CCCAACCTGCGGGCAGGCAGCTGCCTCCACGATGTTTGGTTTTAATAAAGTCTCTTTGTTTCCTTCTCTG
TCTCTCTGCCCCTCTGCTTTCCCTTCTCTGTCCAGTCCCTCTCTCCCCGGCTTCTCCCTCCTTCTAGGCC
TCTGCGGCCCCTCCCCGCCTTCCCTCCCGGGCTGGGGGCGGCGCCGTGACGCGGCGGGGCGGGGGAAGGA
AGGGGGTGTCGCTGACGCCGCGGCGGCCTCCGGCGGCTCCGGCCTTTTGTGCGGGCGGTTGGGTCGGGTG
GGGGCGGCGGCCGCGGAAGGCCAGGCCGGTGCCCTGCGGGGACGCCCAGCGCAGCCCAGCCCCGCGCAGC
CCAGCCCTGCCCTGCCCTGCCCTGCCCTGCGCCCGGGGCGCGCCCACCGCGCCGCATCCATGTTC
GACACCACACCCCACTCTGGCCGGAGCACGCCAAGCAGCTCCCCATCGCTCCGGAAACGGCTGCAGCTCC
TGCCCCCAAGCCGGCCCCCACCTGAGCCAGAACCAGGCACCATGGTGGAGAAGGGATCAGATAGCTCCTC
AGAGAAGGGTGGGGTGCCTGGGACCCCCAGCACCCAGAGCCTAGGCAGCCGGAACTTCATCCGCAACAGC
AAGAAGATGCAGAGCTGGTACAGTATGCTGAGCCCCACTTATAAGCAGCGTAATGAGGACTTCCGGAAAC
TGTTCAGCAAACTCCCCCGAAGCAGAACGCCTCATTGTGGATTACTCCTGCGCCCTGCAGCGTGAGATCCT
GCTCCAGGGCCGCCTCTACCTCTCTGAGAACTGGATCTGCTTCTACAGCAACATCTTCCGCTGGGAGACC
ACGATCTCCATCCAGCTGAAGGAAGTGACATGTCTGAAGAAGGAAAAGACGGCCAAGCTGATCCCCAACG
CCATCCAGATCTGCACGGAGAGCGAGAAGCATTTCTTCACTTCCTTTGGGGCCCGTGACCGCTGCTTCCT
CCTCATCTTCCGCCTCTGGCAGAATGCACTGCTTGAAAAGACGCTGAGTCCCCGCGAGCTCTGGCACCTG
GTGCATCAGTGCTACGGCTCAGAGCTGGGCCTCACCAGTGAGGATGAGGACTATGTCTCCCCCTTGCAGC
TGAACGGTCTGGGGACCCCCAAGGAAGTGGGAGATGTGATCGCCCTGAGCGACATCACCTCCTCGGGGGC
AGCTGACCGCAGCCAGGAGCCAAGCCCAGTGGGTTCGCGCCGTGGCCATGTCACGCCCAACCTTTCCCGA
GCCAGCAGCGACGCAGACCATGGGGCAGAGGAGGACAAGGAGGAGCAGGTAGACAGCCAGCCAGACGCCT
CCTCCAGCCAGACAGTGACCCCGGTGGCTGAACCCCCGAGCACAGAGCCCACCCAGCCTGACGGGCCCAC
CACCCTGGGCCCCTTGGATCTGCTGCCCAGTGAGGAGCTATTGACAGACACAAGTAACTCCTCTTCATCC
ACTGGGGAGGAAGGTGAGGCAGGCGGGCCCAATTCATTCGCCTCCGGTACTTGCAAGCCTCGCTCAGTCT
TAAGCAAGAGGGGATGGATTCGCCCGCAGCACTGAGAATCCAGGGGCAGGCGGGATGGCGTTCAGGCGCT
GTTGCTAGAAATCTCTGTCTTTACTCTGTTTTGAAGGCAGCATGGCAGGGTGAACACAAGCACAGACTGA
AGGCAGCTTGCCGCGGTTCACATCCTGGTCACACCACTTCCTGACCATGTCACGTGGGCAAATTATGTAA
CTTCTCTGAGCTTGCTTTCCTCCTCTGAAAATGGGGTTAACAATGATAATACCCACTTCATAGGATTGTT
GTTAAGAGACTATAAATGAGGCTGGGTGTGATGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCAAG
GTAGGAAGGATTGCTAGAGGCCAAGAGTTGGAGAACAGCCTGGGCAGCAAAGCAAGACCCCGTTTTGTAC

AAAAAAGAAAAAAAAGAAAAAACTAGGTGTGGTGGTGCACACCTATAGTACCAGCTACTTGGCAGACCAA
GGCAAGAGGATCACATGAGGCCAAGAGTTTGAGACAAGCCTGGGCAACATAGTGAGACCCCTGTCTTTAC
AAAAAATTTAAAAATTAGCCAGGCATGGGGTCATGCACCCTATGTAGTCTCAGCTACTGGGGAGGCTGAG
GCAGGAGGATCGCTTGAGCCCAGGAGTTGGAGGCTGCAGTGAGCTATGATTGCACCACTGCACTTCGGCC
TGAGCAACAGAGCGAGACCCTGTCTCTAAAACGAACAAAGGGAGAGAATTAAGTGATTTCATACATTTAA
AGCACTTGGGCCAGTTAGTAAGCATTGACTATTAGTATTTCTATTATCATTATTGTATTCATTTTCTCTC
TCCACTAATTTTCAGCTAGCCTCTATTCTTGTCTCTTATGGGGCAGAGATGGCCTCAGCCCTGCCCCAGA
CCAGCCCAATGAGAGAGAGCAGCCTATTTTTCCTAGTGTCTACAGCGGACACCCTAGGGTTGGCTCTGAT
TGGGCCAACTTGAGAGTGTGCCTGCCCCTGAACCAATCACTGTGGTTGATCGGATAGACCTGGGCCACCC
ACTCACCCCTGAACCAATTGCTGTGCCCCTAGGTTGATGCTCCCAGGAGGCGTGCTATCTGGTAATGAAT
CCCTGTGATTTAAATCTCAGGGCCTAGGTGTGTTCTCCACCCAGAACCAAGCACAGTGGCTGTCTGTCTG
GGCCTGGAGCACGTGCCCACCCTCAACCCATCCCTGAGTCACAGGAGGGCTCTGTGTCAGTGTCCAGCTG
TGCTCCAGGAGGTGGGGTCCGCCCATGCACACTACAAGGCCTGGGAGTGGGGAGGCGTGGCCCGCAGAGG
GAGGGAGGAGTGGAGGCAGTCATCATCCTCGGCTCTCCTCTGTCTCCAGCGGGACTTGGCTGCCCTGCTT
CCCGACCTCTCCGGCCGCCTCCTCATCAACTCTGTCTTCCATGTGGGCGCAGTGTCTTCTGAGATCCGCT
ACCGAAAGCAGCCGTGGAGCCTGGTGAAGTCGCTCATTGAGAAGAACTCGTGGAGCGGCATTGAAGACTA
TTTCCACCATCTGGAGCGAGAGCTCGCCAAGGCTGAGAAGCTGTCTCTGGAGGAAGGCGGGAAGGATGCC
CGGGGCTTGCTATCCGGCCTGCGGCGGCGGAAGCGGCCCCTGAGCTGGCGGGCTCACGGGGACGGGCCCC
AGCACCCAGATCCTGACCCCTGTGCCCGGGCCGGCATTCACACCTCGGGCTCCCTCAGCTCCCGCTTCTC
CGAACCATCTGTGGACCAGGGCCCCGGGGCAGGCATCCCCAGTGCCCTGGTTCTCATCAGCATTGTCCTT
ATCATCCTCATCGCCCTCAACGTCCTGCTCTTCTACCGCCTCTGGTCCCTGGAAAGGACAGCCCACACCT
TTGAGTCCTGGCACAGCCTGGCCCTGGCCAAGGGCAAGTTCCCCCAGACGGCCACAGAGTGGGCCGAGAT
CCTGGCGCTGCAGAAGCAATTCCACAGCGTGGAGGTGCACAAGTGGAGGCAGATCCTGCGGGCCTCCGTG
GAGCTCCTGGATGAGATGAAGTTCTCGCTGGAGAAGCTGCACCAAGGCATCACAGTCTCAGACCCTCCCT
TTGACACCCAGCCCCGGCCCGATGACAGCTTTTCCTGAGGACCCCGGCCACGCAGCTGTTCCCCCACATG
GACAGATGGACACACAGAGCCTCGGCGGCCACTGCTGGCACGGTGTGAGCGCCAGGCATCTCCCACCCGC
CCCTCCCGACGGCCCAACCAGGGGCTGTGCAGACGTGGGGACCACGGAACCGAGATGCACTTTAGACCAG
GGAGCTGGCCCGGCCTCTGGCAGGCCCCCCACTAACTTATTTTGCCCGGCTGAGGTTGTGGGGGGCGCCT
CCTGGGGTGCACGATTCCCTCAGCTCTGGGTTTAATGTATTATATTTATTTGGGGCCGACAGTGCCCCAA
TAAAGGGTCAGAAGTG

>DEX0356_003.nt.3
GTCCAGGAGCTTCTGGCCAGGGCATAGTAGTTGCTAGGGACCCTTTTCTAGGTACTAGGGGAGGGGGCTC
CCCAACCTGCGGGCAGGCAGCTGCCTCCACGATGTTTGGTTTTAATAAAGTCTCTTTGTTTCCTTCTCTG
TCTCTCTGCCCCTCTGCTTTCCCTTCTCTGTCCAGTCCCTCTCTCCCCGGCTTCTCCCTCCTTCTAGGCC
TCTGCGGCCCCTCCCCGCCTTCCCTCCCGGGCTGGGGGCGGCGCCGTGACGCGGCGGGGCGGGGGAAGGA
AGGGGGTGTCGCTGACGCCGCGGCGGCCTCCGGCGGCTCCGGCCTTTTGTGCGGGCGGTTGGGTCGGGTG
GGGGCGGCGGCCGCGGAAGGCCAGGCCGGTGCCCTGCGGGGACGCCCAGCGCAGCCCAGCCCCGCGCAGC
CCAGCCCTGCCCTGCCCTGCCCTGCCCTGCGCCCGGGGCGCGCCCACCGCGCCGCATCCATGTTC
GACACCACACCCCACTCTGGCCGGAGCACGCCAAGCAGCTCCCCATCGCTCCGGAAACGGCTGCAGCTCC
TGCCCCCAAGCCGGCCCCCACCTGAGCCAGAACCAGGCACCATGGTGGAGAAGGGATCAGATAGCTCCTC
AGAGAAGGGTGGGGTGCCTGGGACCCCCAGCACCCAGAGCCTAGGCAGCCGGAACTTCATCCGCAACAGC
AAGAAGATGCAGAGCTGGTACAGTATGCTGAGCCCCACTTATAAGCAGCGTAATGAGGACTTCCGGAAAC
TGTTCAGCAAACTCCCCGAAGCAGAACGCCTCATTGTGGATTACTCCTGCGCCCTGCAGCGTGAGATCCT
GCTCCAGGGCCGCCTCTACCTCTCTGAGAACTGGATCTGCTTCTACAGCAACATCTTCCGCTGGGAGACC
ACGATCTCCATCCAGCTGAAGGAAGTGACATGTCTGAAGAAGGAAAAGACGGCCAAGCTGATCCCCAACG
CCATCCAGATCTGCACGGAGAGCGAGAAGCATTTCTTCACTTCCTTTGGGGCCCGTGACCGCTGCTTCCT
CCTCATCTTCCGCCTCTGGCAGAATGCACTGCTTGAAAAGACGCTGAGTCCCCGCGAGCTCTGGCACCTG
GTGCATCAGTGCTACGGCTCAGAGCTGGGCCTCACCAGTGAGGATGAGGACTATGTCTCCCCCTTGCAGC
TGAACGGTCTGGGGACCCCCAAGGAAGTGGGAGATGTGATCGCCCTGAGCGACATCACCTCCTCGGGGGC
AGCTGACCGCAGCCAGGAGCCAAGCCCAGTGGGTTCGCGCCGTGGCCATGTCACGCCCAACCTTTCCCGA
GCCAGCAGCGACGCAGACCATGGGGCAGAGGAGGACAAGGAGGAGCAGGTAGACAGCCAGCCAGACGCCT
CCTCCAGCCAGACAGTGACCCCGGTGGCTGAACCCCCGAGCACAGAGCCCACCCAGCCTGACGGGCCCAC
CACCCTGGGCCCCTTGGATCTGCTGCCCAGTGAGGAGCTATTGACAGACACAAGTAACTCCTCTTCATCC
ACTGGGGAGGAAGGACTTGGCTGCCCTGCTTCCCGACCTCTCCGGCCGCCTCCTCATCAACTCTGTCTTC
CATGTGGGCGCTGAGCGGCTCCAGCAGATGCTCTTCTCGGACTCGCCCTTCCTCCAGGGCTTCCTACAGC
AGTGCAAGTTCACAGACGTGACCCTGAGCCCCTGGAGTGGGGACAGCAAGTGCCACCAGCGCCGGGTGCT
GACGTACACCATCCCCATCAGCAACCCACTGGGCCCCAAGAGCGCCTCCGTGGTGGAGACACAGACGCTG
TTCCGGCGCGGCCCCCAGGCCGGCGGGTGTGTGGTGGACTCCGAGGTGCTGACGCAGGGCATCCCCTACC
AGGACTACTTCTACACTGCCCACCGCTACTGCATCCTGGGTCTGGCCCGGAACAAGGCGCGGCTCCGAGT

GTCTTCTGAGATCCGCTACCGAAAGCAGCCGTGGAGCCTGGTGAAGTCGCTCATTGAGAAGAACTCGTGG
AGCGGCATTGAAGACTATTTCCACCATCTGGAGCGAGAGCTCGCCAAGGCTGAGAAGCTGTCTCTGGAGG
AAGGCGGGAAGGATGCCCGGGGCTTGCTATCCGGCCTGCGGCGGCGGAAGCGGCCCCTGAGCTGGCGGGC
TCACGGGGACGGGCCCCAGCACCCAGATCCTGACCCCTGTGCCCGGGCCGGCATTCACACCTCGGGCTCC
CTCAGCTCCCGCTTCTCCGAACCATCTGTGGACCAGGGCCCCGGGGCAGGCATCCCCAGTGCCCTGGTTC
TCATCAGCATTGTCCTTATCATCCTCATCGCCCTCAACGTCCTGCTCTTCTACCGCCTCTGGTCCCTGGA
AAGGACAGCCCACACCTTTGAGTCCTGGCACAGCCTGGCCCTGGCCAAGGGCCGGCCTCTGGCAGGCCCC
CCACTAACTTATTTTGCCCGGCTGAGGTTGTGGGGGGCGCCTCCTGGGGTGCACGATTCCCTCAGCTCTG
GGTTTAATGTATTATATTTATTTGGGGCCGACAGTGCCCCAATAAAGGGTCAGAAGTG

>DEX0356_004.nt.1
CGGCCGCCCGGGCAGGTAGAAAAGCAGCCTCCTAGTCAGACAAGCCAGCCGAGAAGAAGGAAGATGAAAG
CCAAATGGAAGATCCTAGTGTGCGGTGCAGTTTTCAACTGACCTCTGGACGCAGAACTTCAGCCATGAAG
GTAACAGGCATCTTTCTTCTCAGTGCCTTGGCCCTGTTGAGTCTATCTGGTAACACTGGAGCTGACTCCC
TGGGAAGAGAGGCCAAATGTTACAATGAACTTAATGGATGCACCAAGATATATGACCCTGTCTGTGGGAC
TGATGGAAATACTTATCCCAATGAATGCGTGTTATGTTTTGAAAATCGGAAACGCCAGACTTCTATCCTC
ATTCAAAAATCTGGGCCTTGCTGAGAACCAAGGTTTTGAAATCCCATCAGGTCACCGCGAGGCCTGACTG
GCCTTATTGTTGAATAAATGTATCTGAATATCCCCCTGTTGTTTCCATTTGCTTTTTCCT

>DEX0356_005.nt.1
GCCCTTGATGTCAAAGTCTCCCACGAAGGAGGCAGGAGCGTTTTGCCCCAGTCCAAGTACTCTGTCCGCC
ACCACATCCCCTACCCCAGTAACGGGGAGCTGCATTCTCAGTATCACGGCTACTACGTGAAGCTGAACGC
CCCCCAGCACCCTCCAGTAGACGTGGAGGACGGCGATGGATCCAGCCAGTCTTCCTCAGCACTGGTGCAC
AAGCCATCCGCTAACAAGTGGAGTCCCTCCAAATCTGTGACAAAACCAGTTGCCAAAGAAAGCAAAYGCT
GAGCCAAAAGCTAAGAAGTCTGAACTTGCTATACCAAAGWAATCCAGCAAGCAACGATTCATGCCCTGCT
TTGGAAAAAGAAGCCAATTCAGGCCCTAATTCAATCATGATTGTCCTTGTCATGCTGTTGAATATCGGGT
TGGCCATTCTTTTTGTTCACTTTCTAACTTGATTGGAATTAGGAAAGTGAAGTCCTGACAACTAGTGGTG
TGTTAGCCCACAGTTCTCTGCAGTGGTGCCGTCAGCCCTTTAAATGGCCCACCCCAGGCAGACTCAGACA
CAGGGAAGGAGGCTTGGAATTAGGATGGGATCGCGAGAGAGATGAAACTTGGAAAATTCAGCCAGAGGAC
CTGTTCGGTTCCCCTGGGAGTGCTTTGTGGCTTTGGGGTCCTCTGGGAGCTGAAGCACACGGCCATGTTG
GAGAGCAAAACTTAATAATAATTCTTTTTTAAAATCTGCTGAAGCAGCCCCATCCGGCGAAGTTCTTGGC
ATTGGCTACTCCATCTGTCCTAGCAGAGTGTGACTAAACTGGAAATGTATGGAGCTGCATTTTTTTTTTT
GATGGAAGTCAACAGCTGCCTTACTATTTTACCATCTGACGTTTTTAGTAGGGAGCTGGGGAAACGACTG
CCCGAGGAATGCGGTCGGAGGATTGTGTTGCTGTGGACGTGGTTCCAACATTCACTCCTATCTCATTAGA
AGAAATAGGTAGCAGCATCACTCACCAGTCTTATGCCATGACCTGCTGCTTTAACATCTCCTGGAATGTT
CTGGGAGAGAACGTTTCGTTTGCTTGTGCACGACTCTGTTCATTTTTGCTGTTTATTTAAATCTGATGTA
TAGTCATCTACCACCTGCGGTCCTGGCCGGACGAAACACACGTGTTGCATCTCAAGGAAAAGAAACAACA
AAATGCTAGTGAAAATGTGCTTATTTTGATAGCAATATACCATCTTTTATGTCATTTATTCTTCCTATAC
TTGTAAAAGGTTATTTATCTTTAAAATTTAGCAATTTGAAGACACTATGCCTTCCTAAGAACTAATTTAA
TTCTAAATATTTTTAACATTACTGAGGTGAATTATTTATGCACTTAGGAAGTGCTAAATTTTAAAAGCTG
AAACACAACAGAATTCTAAGAAATATAGTCCAAACGTTGCATGGATTGCAGTAATCAGTGTTTAAAGGAT
TCAGTTTCTTTGCTGACGTACTTTACAACCAAATAAAATCTTGTCGGTGGCTGTGTTAATTCCCATGAAA
GTTAAGCAAGATGCTATTAATAAACTGCTCTGCTCTTTCTTGTTTTCTTTTTCCAACTTAAATTTCTGTT
GAATACATTCAGGTAGAACATAAAGCCTTGTTCAATCACTGCCTCTCAGTTTTCTGCCTTTCCTGTTTTT
CAAAGTCCTTTTGTAATTTCATTTCACATTTTAAAGGCTTAGACATTTTTATTTTAAATATGTGTCTTTT
ACAGTCTTTTGTCATTCTGACATTTCTGGATTTTTGCTGTTTTATAATTTACCCTTTGTTATTCAGAAGC
ATGCTTACTTATAGAAACTAAATGGTCTTTATAAAAGTAATTACTTAAAAAGAAATCAGGGGAAGAAAGA
TATCTATCTAATCTATTAAATCTTTATAAAACATTACATTGCAGAGGGGGAGCTACTCCTAAATATTTTC
ATGATTTGCATGGTTTAATCAGATTTTTTTTTTTTTACACCATATTAGCTACCTTTTCAATGGAGAAGA
GACAGTTCACACAATTCCCTGGTTAGCACAGATGTGGACTGAGTGCTTTGTCACCTGCAGGGTAGTAACC
CAGTGATGTTTCTTACAGAAGCACAATATGTTGAAAATCCTGGGTGTGACCAATATGGAATAAAGAAGAA
GGCAGAAAGAGAGCAAATGAAAAATTTCAACTTGTATATTCATTTTTTACATTTTGCTTTGACTTTTAAA
TTTAGGAAGTCCGTTTTTACCTGAGAACAAATGTTTAAAGTTCCTGCGTCACTCTCAGTACTCTCACTGC
CCCTCCCAAACCCTATAGCTCCTTACGCTGGGAAAGCTGGTTTTTTAAAAAAATAATAATAAAATATTTA
ATCTTATTAAGTGTTCATTTAAAATGCGTAATGCTTTGGAAATAATGGGTAACAGATAGCGAGAGGATAT
GTTTATAAAGTGAGCATGTTGGTCCCATTTATAAATATATGTATGATTTATAAGCTTTTTTAAAACAAAG
CTCAAATTGTTGGTATTTTTCTAAAATGTGCACAGCTGTATTTTACATGAAGGCTCTTTCTAATGGGTTG
TTATACTGTACTCAACATTTTGGACAGCACATGAAGTCTGCCAATGTACTTAATAAAACATGACTTTGTT
TATTTAAAGTTTCTTGCTGTGAAAAAGAACTCCCTACCTGTGAGTTCCTTTATTTATAATTCTTGAAACC
AAAATGTATAATGTACAGTTTTCACAACTGTATCTGCTCTAATAAAAAAAAGTTGGTTATTAAAAAAAAA

AAA

>DEX0356_006.nt.1
GATTCTCTTAATTTCCACCGCTACAGATGTACATGTACATCCAGACAACAACACTTACTGTCTCCATGAG
TTTAAGTGCTTCAGTATCTCTGGGCATGCTCTATATGCCCAAGGTTTATATTATAATTTTTCATCCAGAA
CAGAATGTTCAAAAACGCAAGAGGAGCTTCAAGGCTGTGGTGACAGCTGCCACCATGCAAAGCAAACTGA
TCCAAAAAGGAAATGACAGACCAAATGGCGAGGTGAAAAGTGAACTCTGTGAGAGTCTTGAAACCAACAC
TTCCTCTACCAAGACAACATATATCAGTTACAGCAATCATTCAATCTGAAACAGGGAAATGGCACAATCT
GAAGAGATGTGGTATATGATCTTAAATGATGAACATGAGACCGCAAAAATTCACTCCTGGAGATCTCCGT
AGACTACAATCAATCAAATCAATAGTCAGTCTTGTAAGGAACAAAAATTAGCCATGAGCCAAAAGTATCA
ATAAACGGGGAGTGAAGAAACCCGTTTTATACAATAAAACCAATGAGTGTCAAGCTAAAGTATTGCTTAT
TCATGAGCAGTTAAAACAAATCACAAAAGGAAAACTAATGTTAGCTCGTGAAAAAAAATGCTGTTGAAAT
AAATAATGTCTGATGTTATTCTTGTATTTTTCTGTGATTGTGAGAACTCCCGTTCCTGTCCCACATTGTT
TAACTTGTATAAGACAATGAGTCTGTTTCTTGTAATGGCTGACCAGATTGAAGCCCTGGGTTGTGCTAAA
AATAAATGCAATGATTGATGCATGCAATTTTTTATACAAATAATTTATTTCTAATAATAAAGGTCGCAGG
CGACCAGATCTTTCTATCCACATTTCCTATCAATACCTCCACAATTATCTA

>DEX0356_007.nt.1
GCTAAAGCACCCAGAGGGACTAGACAGACTCTGGCTAGAGGCACTTGTCCTGAAATCTGATTCCTGCCTG
ATGATGGTGTCTCTTGTATCAATCTCTCTCTCTGGTCTTTGAATTGGATAGAAATTATAACCCAGCTTCA
TCGTTCTCCAAGTTTGGAAATCCTTCTGTGTGAGTATTAAGGGCATCACTTTGAAGAAAGGACAAATGTT
CTCGATGGTTGATTTGGAAGAGTACCTTTATGGTAGTGTTTCTCAGTGGGGACAATTTGTCCTCCCCTTC
CCACTCTGGGATAATCTGGCAATGTCTGGAGACATTCTGGGTTCTCACAAATAGGAGGAGGATGCCACTG
GCATCTAGTAGGTCAGCGATGCTGCTAAGCATCCTGCAGTGCCCAGGACAGCCCCACAAAGACTCATCCA
GCCTGAATGTCAATAGTGCTGAGGTCGAGGAGCCTTACTTTACGGGAACTAGGCAGGAGGAGTCTGTATG
CCCAAGGACGTTGGTCATTATCACGTGTCCACTTCAGTACAGCACTACTCAGAGGCAATGCTTTTTGGGT
AACGGATTTTTCTCTTAACAGGTTCTTCATTTTGAAAATGTCAAAGATGTACCCTTTGGATTTCAAACAG
TAACATCGGATGTAAACAAACTTAGTTCCTTTTACTCACTGAAACTAATCAAGCGGCTCTACGTAGACAA
ATCTCTGAATCTTTCTACAGAGTTCATCAGCTCTACGAAGAGACCCTATGCAAAGGAATTGGAAACTGTT
GACTTCAAAGATAAATTGGAAGAAACGAAAGGTCAGATCAACAACTCAATTAAGGATCTCACAGATGGCC
ACTTTGAGAACATTTTAGCTGACAACAGTGTGAACGACCAGACCAAAATCCTTGTGGTTAATGCTGCCTA
CTTTGTTGGCAAGTGGATGAAGAAATTTCCTGAATCAGAAACAAAAGAATGTCCTTTCAGAGTCAACAAG
ACAGACACCAAACCAGTGCAGATGATGAACATGGAGGCCACGTTCTGTATGGGAAACATTGACAGTATCA
ATTGTAAGATCATAGAGCTTCCTTTTCAAAATAAGCATCTCAGCATGTTCATCCTACTACCCAAGGATGT
GGAGGATGAGTCCACAGGCTTGGAGAAGATTGAAAAACAACTCAACTCAGAGTCACTGTCACAGTGGACT
AATCCCAGCACCATGGCCAATGCCAAGGTCAAACTCTCCATTCCAAAATTTAAGGTGGAAAAGATGATTG
ATCCCAAGGCTTGTCTGGAAAATCTAGGGCTGAAACATATCTTCAGCGAAGACACATCTGATTTCTCTGG
AATGTCAGAGACCAAGGGAGTGGCCCTATCAAATGTTATCCACAAAGTGTGCTTAGAAATAACTGAAGAT
GGTGGGGATTCCATAGAGGTGCCAGGAGCACGGATCCTGCAGCACAAGGATGAATTGAATGCTGACCATC
CCTTTATTTACATCATCAGGCACAACAAAACTCGAAACATCATTTTCTTTGGCAAATTCTGTTCTCCTTA
AGTGGCATAGCCCATGTTAAGTCCTCCCTGACTTTTCTGTGGATGCCGATTTCTGTAAACTCTGCATCCA
GAGATTCATTTTCTAGATACAATAAATTGCTAATGTTGCTGGATCAGGAAGCCGCCAGTACTTGTCATAT
GTAGCCTTCACACAGATAGACCTTTTTTTTTTTTTCCAATTCTATCTTTTGTTTCCTTTTTTCCCATAAGA
CAATGACATACGCTTTTAATGAAAAGGAATCACGTTAGAGGAAAAATATTTATTCATTATTTGTCAAATT
GTCCGGGGTAGTTGGCAGAAATACAGTCTTCCACAAAGAAAATTCCTATAAGGAAGATTTGGAAGCTCTT
CTTCCCAGCACTATGCTTTCCTTCTTTGGGATAGAGAATGTTCCAGACATTCTCGCTTCCCTGAAAGACT
GAAGAAAGTGTAGTGCATGGGACCCACGAAACTGCCCTGGCTCCAGTGAAACTTGGGCACATGCTCAGGC
TACTATAGGTCCAGAAGTCCTTATGTTAAGCCCTGGCAGGCAGGTGTTTATTAAAATTCTGAATTTTGGG
GATTTTCAAAAGATAATATTTTACATACACTGTATGTTATAGAACTTCATGGATCAGATCTGGGGCAGCA
CCCTATAAATCAACACCTTAATATGCTGCAACAAAATGTAGAATATTCAGACAAAATGGATACATAAAGA
CTAAGTAGCCCATAAGGGGTCAAATTTTGCTGCCAAATGCGTATGCCACCAACTTACAAAAACACTTCGT
TCGCAGAGCTTTTCAGATTGTGGAATGTTGGATAAGGAATTATAGACCTCTAGTAGCTGAAATGCAAGAC
CCCAAGAGGAAGTTCAGATCTTAATATAAATTCACTTTCATTTTTGATAGCTGTCCCATCTGGTCATTTG
GTTGGCACTAGACTGGTGGCAGGGGCTTCTAGCTGACTCGCACAGGGATTCTCACAATAGCCGATATCAG
AATTTGTGTTGAAGGAACTTGTCTCTTCATCTAATATGATAGCGGGAAAAGGAGAGGAAACTACTGCCTT
TAGAAAATATAAGTAAAGTGATTAAAGTGCTCACGTTACCTTGACACATAGTTTTTCAGTCTATGGGTTT
AGTTACTTTAGATGGCAAGCATGTAACTTATATTAATAGTAATTTGTAAAGTTGGTTGGATAAGCTATCC
ATGTTGCAGGTTCATGGATTACTTCTCTATAAAAAATATGTATTTACCAAAAAATTTTGTGACATTCCTT
CTCCCATCTCTTCCTTGACATGCATTGTAAATAGGTTCTTCTTGTTCTGAGATTCAATATTGAATTTCTC
CTATGCTATTGACAATAAAATATTATTGAACTACA

>DEX0356_008.nt.1
TAAAAGAAGATGGGGCCTTAACATTTTTGTAAGGTTAGGTGAAAAGGGTGAATGTAGAGGGGCTTGACTG
AGAAGAGTGGAAGATGTAGTGACTTAAGTGAACCAAAGGAAGAATACCCTTTCCTCAGAAGCAAGGAGGA
AGGACCTGAGAATAGTTAGGGATATATGTATACCTTCTTGGAAGAGGGAAGGACTTTAGCCATTTTGTTT
CAGTATTTTTACCTAAATGTATTTCTCTCAAAGGTCGAGGTCAAGGAGACATTCTCATAGACGTTACACT
CGATCCAGATCCCACTCTCACTCTCATAGGAGACGATCTCGAAGTAGATCATATACACCAGAATACCGGC
GGCGAAGGAGCCGAAGCCATTCTCCAATGTCTAACCGGAGAAGACATACTGGCAGCAGGGCAAATCCAGA
TCCCAACACTTGCCTTGGAGTGTTTGGCCTCAGTTTGTACACAACAGAGAGGGATCTTCGTGAAGTATTT
TCTCGATATGGACCATTGAGTGGTGTCAATGTGGTTTATGATCAGCGAACTGGGCGATCTCGAGGATTTG
CTTTTGTGTATTTTGAGAGAATAGATGACTCAAAGGAGGCTATGGAAAGGGCAAATGGAATGGAGCTGGA
TGGTAGAAGAATTCGGGTGGATTATTCTATAACCAAGAGAGCGCACACACCAACACCAGGCATCTACATG
GGCAGACCAACTCATAGTGGTGGGGGTGGTGGAGGAGGCGGCGGCGGTGGAGGTGGAGGTGGTGGCAGAC
GTCGAGATTCTTACTATGATAGAGGATATGATCGTGGGTATGACAGATATGAAGACTATGATTACCGATA
CAGAAGACGATCACCTTCTCCTTATTATAGTCGATATAGATCACGATCAAGATCTCGTTCCTACAGCCCA
AGACGCTATTGATAACGGAATGGTTGCAATTAAGGACATTTTTTTTCCTCTTTTTTTTTTTTTTTTTTTT
AATTCTGAGATTTCCCCAAGCTGTGGATTCTTCCTACTCCTTAAGAAAAAAACTTTGGTTTATTTAGCAT
CTACACTTTTGTCAGTTGTGTTGCTGTTTTCCACCCATTTTATTATACTCTTAAAAGATGTAATTGTTGT
CATTTTGAACAGTTAAACATCTTGAGTATAAAAAGAACCCCAATGTTATGTTATGCTTTGTAAATTTTTT
TTTTTGCTTTTACCTAGATAAACTTCTAGCTAATCAAATAAGGAAAGAAACTGTCTTTTTAAAGCTTCTT
TTGTGTTAGATACTGTATTAGAGATCTGCATTTATCATGAGTTCCTTTTTTTTTTTAACTTTATTTTTGG
GAAAGTAACACATGAAGTAGTTCAGTCATGTCAGGTTTGTCTGGGGTGGAATGGAACAGTCAGGTAGTTG
AAAGTTTTTTTTTAGAGATGAAAAGCTTGTGAACTCCTGTAAAACATGCTGTATTTGAAATACATCTGTT
AAAACTTAAAAACTAAAGTGTGATTTTTTTTGTTGTTGTTAAGTTTATTGAAGTTTATTCCCTTAATCAG
TGAAGGACAACCCTTATTTATTACCTAGAGCAGTTGTATAATTCGCTGTTAGAAATTTTGGTATTGGGAC
AGTGGATAAGCAGGTTATCACATAGCATAGAATTCTTAAAAGATACATGTGGGGAAAAGTAGTCTCAAAT
AAAAGCTAATTTCTTTGAAAAAAAACACACAACCCACACGCGCACCGAAACAACCCAGCGACCAACAGCA
GTCGCCACTCCGGGCGGGGGACCCACATAAGCGATCCCCCGCCGGAGGCGCACCAAGACAGAGCCGCACC
CACGTAGTGGCGCCACGAAGGTGGCCACTTAAGCAGCGCCTAATTCAGACGCACGCCAGCCCGCGACAAA
ACGAGCCCCGATGTAAAGGATCCTCCGCAGGCACGTTGCGAGCCCACCTGCGCCTGAGGGGCACAAGTCG
CGACATCGACGCAGCAAAAACCACGAGCAATCAGTCACAGAGCATCAGGCTACAAACGCGATTGCACGCG
TCAACGGCCCAGATGAAAGAAACA

>DEX0356_009.nt.1
AATTTATAAGATTCATAGAAATAGACAATGCAAAGAGTTCCCTGAGGACAGGAACCACATCTGTATTGTT
CACTTCTGTAACCCAGTGCTTTGCCCATATTAGATGCCCAGAAAATAACAATTGAGTTTATGAATGACCT
CTGATATCTCTACATCCTTAATTCAGTCTTATTTTTATCCACTAAAATCATGAGATTAGTGAAATTGTTT
AAGCACGTAAGGATAGTTTTACTGTTAATTTAGATTTTACCAATAAGCAAGAGTAGTACTATTGCTGGAT
TTATATGGGTTTGATATAATTTTATAATCCAAAAGAAAGATTGGAAAGGGATGACAAGTTGGTTAATTCA
TTCTTAACTGAAAGATAGTGGATAGCATAGCACATAAAAGCACTGGACTCTGGAGTAAAACTGCCTGTAT
TTCAGTCTCGGCTATGTGATCTTGGTCAAGTTACCTTACCTCTTTGTGTATTAGCTTTCACATGTAAAAT
GGGAATGGCAGTAATAACACTTCCTTATAAGGTATAATATTGTACATTTTGTGAGGATTAAATGAGTTAA
TACAAATATAACACTATATTATTATATTGATGTTGTTAGATGTTATTATTGAAATTGCTAAATTATTTGA
TAGTTTTACCCAAAGTTGAGATTAGTCTTGATCACTTATAAATAATTTTTATCTTCATGATTAGGCTTTT
CAGAATTCTTTAGCTCATCATTAATGACTGTTATTCTCATAGCATTGGAGGCGTTGTACTGGTAAACCTG
GCAGGGTCTGAAAAACCTGCTGGAAGAGACACAGTAGGTTCCATTTGGTCTCTTGCTGGAGCCATGCTCT
ATGCTGTCTATATTGTTATGATTAAGAGAAAAGTAGATAGAGAAGACAAGTTGGATATTCCAATGTTCTT
TGGTTTTGTAGGTTTGTTTAATCTGCTGCTCTTATGGCCAGGTTTCTTTTTACTTCATTATACTGGATTT
GAGGACTTCGAGTTTCCCAATAAAGTAGTATTAATGTGCATTATCATTAATGGCCTTATTGGAACAGTAC
TCTCAGAGTTCCTGTGGTTGTGGGGCTGCTTTCTTACCTCATCATTGATAGGCACACTTGCACTAAGCCT
TACAATACCTCTGTCCATAATAGCTGACATGTGTATGCAAAAGGTGCAGTTTTCTTGGTTATTTTTTGCA
GGAGCTATCCCTGTATTTTTTTCATTTTTTATTGTAACTCTCCTATGCCATTATAATAATTGGGATCCTG
TGATGGTGGGAATCAGAAGAATATTTGCTTTTATATGCAGAAAACATCGAATTCAGAGAGTTCCAGAAGA
CAGCGAACAGTGTGAGAGTCTCATTTCTATGCACAGTGTTTCTCAGGAGGATGGAGCTAGTTAGCTGTCT
GTTGTCTGTAGCCCAGCTTGATAATGGAACTATACAGCGAAGAGACAATCTCTGGCAAGTTTTTGTAGAA
AAAATGTTTCAGTGCCTAGTCTGAAAAATAACAGTTTGAGTTCTTTGAAACTCTAAAATATATTTTTCTC
ATACCTGTTTTCTTCATTTTCATAATGAAGCACTTTGCTATGTAGCTGTGTACATATCACTACAGTTATA
GGAAGTTTCAGTCTACAGTCCATCCAAAGGACCAACCTGCCTTACACATCTCAAGGAATTCAGCTGTTGA
AATCATTTGAACTAATCAAGGAATAAATCCTAATGTTCTGGGACTTTATTTTCACATGTTAAATGCTGGA
ATATATTATGAAAATGTTTTCAAGAAATCACTTAAGTGTTCATAGACCAGTATTTCTGACAGGTAAAATG

CTAAAATAAGCTACCTGTAATAAGTGTGGATTATATTTTTGGGTTTTGTAGAATATTGCAAATTAACCAC
ACAAAAAATGTTTAATTTATGCAACAAGCATGTTTGTGCAAATTTCATGGGACTTTAAAAAGAATAAGTA
TTTGAGAAAATATCTGGTTCACTTACACTACATTTACTGTATTATTCTTTTATAGCATTAGGTGCCTTGT
ATTTTAAATCTGTGACAAACCATGGCAAATTTTTAAAGGGGAAGTATTATTATAAAATGAAGAAATATGT
ATTTCTAAAGGCTATATTGCTGTAAACTTAATTGATAAAGCTCTGTTTAATTTAGAGTTTTGAAGAAATA
GTCTCCCTTCAATTAAGAAATTTTCATAATGGAATGATTTAAATTGAAGTGACAAAGAGTATTATTAAAA
TACAATGTTTATACGTCTATTTGTGTATTGTAGATGTATCAAGTGATTTCTAATTTTTTTCACATATGAA
TGTGCCAGATTACTCTAGAACTAGATGTCTCTTCTTTAAATAATTTTAGTTTTCCTGAATAAATTTGTAA
TGGTTAAAGTACCAAGTAAGTAAGGCGAGAAGGGATTCTGTTTTTAAAATCACATCAGAACTTTTCCTCT
ACTAAGATTATAAATTAAATGTAAAATACTCCTAATTGCAATTCTTAAACTTAGGCCTTACATGTACTTA
TTATGCAACTGCTCCTGGACTCTATTCACCATAGATATCAGTAAACGTATGTCCCAGGATTCACAGGCTT
TTGATTAATCAATATTCTTTTCAAGTTTGCTGTGAAGAGTTTAGTTCTCTTCAAAATTTCTTAACTAATC
TGATTTTTAAGAATTCTCTTTGCAGTGTTTAGCTTCCTATTCACATTCTTAAAATTGCTTTGGTGTTACC
ATGAGTCTAAAATGAAGTTTAGCCTTCCTTTTGTTTCATTCTGAGAACTTCTATATTATATTACCTTTAA
AAATTGTTTATGATATTAAATTTAAAATACAAACAGCTCTCTTTTTTTTTTTTTTTTTTTTAATCATCC
AGCCCAAAGTGGCAAAAACAGCTCTTTTCTCATTTGGCACCACCAATAACGCAAGTTAAAAATAATGTTG
AGTTTATTATACTTTTGACCTGTTTAGCTCAACAGGGTGAAGGCATGTAAAGAATGTGGACTTCTGAGGA
ATTTTCTTTTAAAAAGAACATAATGAAGTAACATTTTAATTACTCAAGGACTACTTTTGGTTGAAGTTTA
TAATCTAGATACCTCTACTTTTTGTTTTTGCTGTTCGACAGTTCACAAAGACCTTCAGCAATTTACAGGG
TAAAATCGTTGAAGTAGTGGAGGTGAAACTGAAATTTAAAATTATTCTGTAAATACTATAGGGAAAGAGG
CTGAGCTTAGAATCTTTTGGTTGTTCATGTGTTCTGTGCTCTTATCATCACACAGGTCATGTGTTGTTAC
TCAGGATTCTGGAAGTACTAAGCCGTAGTTAACAGGCTGGATAGATCTTCAGCCAATCTTCTTTCATTAT
ACTGCTGCTTTTCTGTTTCTTAAAAAACAAAAATCAATGAACAACTTCTTTAGAAGGAAGCACCACTGTT
CAATTGGTTAACTGAAAGTATGAATCACTGCCGCTTTGCTTGCCCACCTGTTTGCTTTTTCTCACAGGTT
TTATTTATCTGAATGATTGTTGCATTTGAATACTGTAGCCATGGTGAGTGAGCAGTTGAAGACTTCCTTT
CTGCACCTTCATAGTTAAGGGTTCATCTTCTCAAGAAATAAAGTTGTGCTGGGTTGTTTCAATTCTGTAG
ATCACCTAGAAACTAGGTAACAAGGCCCTTGGATCACAATGAGCACATTGCAGGTCTTTGGAAATGCTGG
GAAGGGGTTACATTCAAGTAACTGCTGACGTCCTTCCTCTCTGTCAGACCCATGTTGGCCTGTAATTATA
TTTTTCTGAATCTAAAAACAAATAGAAATTTCTGATATCTTTTCAGTCTCCTTCTTTTCTTTCCATTCAT
CACTTAAATCTCATTATTGTATAACGTTTCAAATTATGACCTGGATTGAAGGAAGTCTGTTTTGTCAGGG
ACTTTGTTAGGTGAACATCAGAATCTCAACTACAACAAAGCTGAAAAATGAGGCAGCCTTAACAGTAAAT
GCTGGCTAAAACCAATTCTTTAGATTGTTTTTGCCACCATTTGCTTACTTGGAATAGCTTTCACATTTAT
TTTTCTTATGAAGAAATCCAGTGTCAGTCAGACCATGTAACATCAAAGCAAAATTTCATTGAAGCCTATC
CTTGTCTTAATTATTCTAAAATAAAGTACAATAAAGTGAATTATTTTGCCCTTTGTAGTAGCATAGTGTT
AAGTTATATATATATTTTAATATGCATACTCCTAAGTAAAGTTTCAGAGACCACATGCTGGAGAGCAGTG
GCTTTCAGACAGCATCAGCATCACCTGAAGGCCTTGTTGAATACACATTATTGGAAGAGAGGGGAGAACA
CAGGCTCATGCCTGTAATTCCAGTGCTTTGGGAGACTGAGGCGGGAGGATCACGTGAGCCCAGGAGTTCA
AGATCAGGCTGGGCAATACAGCAAGACCCCTGTCTCTACAAAAACAAAAACAAAAAACAAACAGGCATGG
TAGTGCATACCTGTAGTCTGAGCTGCTTGGGAGTCTGAGGCAGGAGGATCACTTGAACCCAGGAGTTCGA
GGCTGCAGTGAGTTATTATTGGGCCACTGCAGTCCAGCCTGGGCAGCAAAGTGAGACCCTGTCTGTTAAA
AAAACAAAACCAAAAACTTGATTGCTGGGGGCCACCCTAGAGTTTCTGATTCAGTAGGTCTGGCGGTGGG
GATGGTGGGCCTGGTAATTTGCCTTTCTAACAAGCGTCCAGGTAATGCCCCAGCTGCTGGTCCCAGGACC
ACACTTTGCAGATTATTACTCTGAGAAGAGCACTAACTCTGACGTTGGATTCCTTCTGTTCCTCATTAGT
GAATTTATTTGGGGGAGATTATTCAATCTGGTAGATCCGCATTTACTTAGCTGTGAAGTGGAGGTTAATG
CAACCTGCCTGAGACTGTGAGAATGAGAAACTCTAATTGAAATACCAGGCCCAGCGTGCCTGGATGCTTG
AAAAATGCTGACTCCCATGGTCCGTCCTTCCCTCACTCTTCTCAGCTATAAAGTGAGGAGCGAAATTCAG
TATCTTTTAAGATTCCTCTCAGGTGTAAAATGCTGTGATTCTAAAAGTTTAATTTTGTAGTTCCAGAGAC
AATGCCAAACACCTGTGTGTGCTTAGTGCTTTGTAAAGAGTGGTAATAACAGTGACACCACATAGGCCGT
AGTTTTGTAGTTCCAGAGACAATGCCAAACACCTGTGTGTACTTAGTGCTTTGTAAAGAGTGAATAACAG
TGACACCACATAGGCCATAGTTTTGTAGTTCCAGAGACAATGCCAAACACCTATGTGTACTTAGTGCTTT
GTAAAGAGTGGTAATAACAGTGACACCACATAGGCCGTAATTTTGTTTTTGCTTTTGTTAGGGTGTTCAG
CATCTTGACTCCTTAAATAAGGTTTTCAGTCAGATTTATAAAGCAGCATTGAAGAAGTTGTATTATATTT
GAGTACAGTTGGAATCCATGGCAGTTTTATGGAGGCCAAATATATCAGATGAATAGGCTAATTTCCCTGA
GTTGCTGGCCTTATTAGTCCCTACTCTGTTGCCCCCTGGTTACCTAATGCATGTTCAAAAGGTATAGTTT
TCATCACCTGCTTGATCTTGATTTTCAGACAGAGCACTTTGGGAGTCCGAGGCGGGTGGATCACCTGAGA
TTAGGAGTTCAAGACCAGTGTGACCAACATGGAGAAACCCCTTCTCTACTAAAAATACAAAAAATTAGCC
AGGCATGGTGGTGCACGCCTGTGGTCCCAGCTACTTGTGAGGCTGAGGTAGGAGAATCGCTTGAACCCAA
GAGGCAGAGGTTGCAGTGGGCTGAGATCACACCACTGCACTCCAGCCTGGGTGACAGAGCGAGACCCTGT
TTCAAAAAAATAAGAAATAAAAAGTGAGTCTTCATATTCAAATTAGTATATGAGGTTAGGCAACAATTCT
TATATGGGACCAGGTATATTTAATTTGCCTTTCTTTGCAGTCATCCTAGTAAATCCTTGTAAAGGGTTAT

TTTATATACAATTTTGTAGTATATTCTAAATCCCCTTTGGCAAATAAAGCTGAAGTTATCTAGTGTCAGC
ATACTGTGATTTGTATATTTATATAAGACAAGATTGCATCATTCTGATTAAGTTGCAGCATTTTAGTTTT
ATCTGATAGGCCTGTATTGAGTAATCAGTAGCTATGTGCAAATATTAAGTCAACTTAATTACCAGACCCA
GAAGGCCAGAGGTAATCAAAGAGTACTTTCGGAAGATGAAATATTTATAATGAGTTATGGAAACAATGTT
TTAGAAATGTTTTTCTTGGTGTAATTTAATGCAGATTTTTACCTCAATTTTGTTAAAAAACTTGATAATT
TGATCCCTCCATCTATTTAATATTTAATATTTTCAGATTCTAGAGTATTTAAAATTTAAACACAATTATA
AAATTAATTATAATTCCTACTTATGTTAGCTATGATTTAGGGCCAAAACTACCCTGGTTTTGTCTTGCAA
CTTGAAACATTGTGTATTTTAAACTATATAATTTGGGAACATTATTTTATAGACTGTGGATAACCAAATA
TAAGATTTGTGATTTTTAATGCTTCATTTGTGTAAAATCATTTGTATATTCTATTTATAGTATGCATTTT
CTAGCTCATTAAAAGTACAGAAAAACCTCAAA

>DEX0356_009.nt.2
GGCGGAGGTTTAGCCACTCCCTCGTCCTTTGCTCTGCCCCAGACACCCACGCTCGCTTTCCTCTTCTTCC
TCCACCTCGCCCTACTTCCCTTCTTCCTATCTTCTCTCTTCTTTCGCCTCGCCGCCGCGAGCCCTCGCTC
GGCGCCCAGAGCCTCGCATTACTTCCCCTAGCTCCGCCAAGCCTCGCTGCATACGCGCCCGCGCGCAGTC
CGCCGTCGCGCCCCTATGACGTCACCCGCGCGCCACCGCTTCCGGGCTGGGGACTGGGGGCGGGGCCGCG
CGGAGGGCAAGCCGGGGGCGGGGCCGGGGCCAGAGCAGATCTCCGGGCCTCCGCGGCCATAGCTGACTGT
GCCGTCCCTTCCCCTCACCCGCTCCACGCCCTCCTGGGCCGAGTGGAGTTGGGTGGTGTCGGGAGCCTCT
CCCTGAGGGGCACCGCGTCTTCAGGAGCTGGGCCTCCAGTGCGGCGCGATGTCAGGCGCGGTGACAGCTC
TGTGAGTCCGAGGCCGCGGCCGTGGCGCTGGGCGGCTGCGGGGCCTGACCGGTCCGCTCATGGTGCCGCC
ACGACGCCATCGCGGGGCAGGAAGGCCAGGGGTGCTGAGTTCTTCACCTCCTTTTAGACTGAGATCTGCC
AAGTTTTCCGGCATTGCTCTTGAGGATCTCAGAAGGGCTCTTAAGACAAGACTGCAAATGGTGTGTGTAT
TTGTCATGAACCGAATGAATTCCCAGAACAGTGGTTTCACTCAGCGCAGGCGAATGGCTCTTGGGATTGT
TATTCTTCTGCTTGTTGATGTGATATGGGTTGCTTCCTCTGAACTTACTTCGTATGTTTTTACCCAGTAC
AACAAACCATTCTTCAGCACCTTTGCAAAAACATCTATGTTTGTTTTGTACCTTTTGGGCTTTATTATTT
GGAAGCCATGGAGACAACAGTGTACAAGAGGACTTCGCGGAAAGCATGCTGCTTTTTTTGCAGATGCTGA
AGGTTACTTTGCTGCTTGCACAACAGATACAACTATGAATAGTTCTTTGAGTGAACCTCTGTATGTGCCT
GTGAAATTCCATGATCTTCCAAGTGAAAAACCTGAGAGCACAAACATTGATACTGAAAAAACCCCCAAAA
AGTCTCGTGTGAGGTTCAGTAATATCATGGAGATTCGACAGCTTCCGTCAAGTCATGCATTGGAAGCAAA
GTTGTCTCGCATGTCATATCCTGTGAAAGAACAAGAATCCATACTGAAAACTGTGGGGAAACTTACTGCA
ACTCAAGTAGCGAAAATTAGCTTTTTTTTTTGCTTTGTGTGGTTTTTGGCAAATTTGTCATATCAAGAAG
CACTTTCAGACACACAAGTTGCTATAGTTAATATTTTATCTTCAACTTCCGGACTTTTTACCTTAATCCT
TGCTGCAGTATTTCCAAGTAACAGTGGAGATAGATTTACCCTTTCTAAACTATTAGCTGTAATTTTAAGC
ATTGGAGGCGTTGTACTGGTAAACCTGGCAGGGTCTGAAAAACCTGCTGGAAGAGACACAGTAGGTTCCA
TTTGGTCTCTTGCTGGAGCCATGCTCTATGCTGTCTATATTGTTATGATTAAGAGAAAAGTAGATAGAGA
AGACAAGTTGGATATTCCAATGTTCTTTGGTTTTGTAGGTTTGTTTAATCTGCTGCTCTTATGGCCAGGT
TTCTTTTTTACTTCATTATACTGGATTTGAGGACTTCGAGTTTCCCAATAAAGTAGTATTAATGTGCATTA
TCATTAATGGCCTTATTGGAACAGTACTCTCAGAGTTCCTGTGGTTGTGGGGCTGCTTTCTTACCTCATC
ATTGATAGGCACACTTGCACTAAGCCTTACAATACCTCTGTCCATAATAGCTGACATGTGTATGCAAAAG
GTGCAGTTTTCTTGGTTATTTTTTGCAGGAGCTATCCCTGTATTTTTTTCATTTTTTATTGTAACTCTCC
TATGCCATTATAATAATTGGGATCCTGTGATGGTGGGAATCAGAAGAATATTTGCTTTTATATGCAGAAA
ACATCGAATTCAGAGATGGAATCTTGCTCTGTCGCCCAAGCTGGAGTGCAGTGGCGAGATCTCGGGTCAC
TGCAACCTCTACCTCCTGGGCTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCG
CACGCCACCACACCCAGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGTCAGGATGG
TCTCGATCTCTTGAACTCGTGATCCGCCCTCCTTGGCTTCCCAAAGTGCTGGGATTATAGGCGTGAGCCA
CCGCGCCCGGCCCACAGCATCATTTTTCAAGGGTCTCTAGGGTGACCAGGTGTCCGGGTTTGCCAAGGAC
AGTCCCAGTTTCTACCTGCTGTCCTAGGACAATCATGAATAGTGCCCCTTTCAGTTTCAAAAGTGTCATA
GTTTAGACATTAAATTACCTGACCACCCTTATATGCCCCATGTTTCTTAACATCTAGGCTGCTGACAAAA
TTCACATTTTCTGACTGTCAGTCTAGTGTTCTCTATACTAGACTGATACGGATCCACAAAGGAAGCAATA
TTGGCCAACAATTAATCTTCAGTCTTTGGATTAGTGAGAGGTGACAACGTGCTGGCAGCCGTCACAGCCC
TCGCTCACTCTCCCCGCCTCCTCTGCCTGGTTTCCCAGTTTGACGGCACTTGAGGAGCCCTTCAGCCCGC
CGCTGCACTGTGGGAGCCCCTTCCTGGGCTGGCCGAGGCCGGAGGCGGCTCCCCCAGCTTGCGGGGAGGT
GTGGAGGGAGAGGCGAGGGCGGGAACAGGGGCTGCGCGCGGTGCTTGCGGGCCAGCGCGAGTTCTGGGTG
GGTGTGGGCTCGGCGGGCCCCGCACTCGGAGCGGCCGGCAGGCCCTGCCGGCCCCGGGCAATGAGGGGCT
TAGCACCCGGGCCAGCGGCTGCGGAGGGTGTGCCGGGTCCCCCAGCAGTGCC

>DEX0356_010.nt.1
AATCACTATAGGCGCATTGGTGCTCTAGATGCTGCTCGAGCGGCGCAGTGTGATGGATGTGGTCGCGGCC
GAGGTGTCTGTTTGACTTTACGCCCATCTCAGGACACTTCCGTAGACTGTTTAGGTTCCCCTGTCAAATA
TCAGTTACCCACTCGGTCCCAGTTTTGTTGCCCCAGAAAGGGATGTTATTATCCTTGGGGGCTCCCAGGG

CAAGGGTTAAGGCCTGAATCATGAGCCTGCTGGAAGCCCAGCCCCTACTGCTGTGAACCCTGGGGCCTGA
CTGCTCAGAACTTGCTGCTGTCTTGTTGCGGATGGATGGAAGGTTGGATGGATGGGTGGATGGCCGTGGA
TGGCCGTGGATGCGCAGTGCCTTGCATACCCAAACCAGGTGGGAGCGTTTTGTTGAGCATGACAGCCTGC
AGCAGGAATATATGTGTGCCTATTTGTGTGGACAAAAATATTTACACTTAGGGTTTGGAGCTATTCAAGA
GGAAATGTCACAGAAGCAGCTAAACCAAGGACTGAGCACCCTCTGGATTCTGAATCTCAAGATGGGGGCA
GGGCTGTGCTTGAAGGCCCTGCTGAGTCATCTGTTAGGGCCTTGGTTCAATAAAGCACTGAGCAAGTTGA
GAAAAACAACAACCACAACACAAAATAAAAGCTCTTTTTTTTCTTCTTTTTTGGGCAAAAAGGCCCTTTT
TTTTCGGGAAAGGGTTT

>DEX0356_010.nt.2
AATCACTATAGGCGCATTGGTGCTCTAGATGCTGCTCGAGCGGCGCAGTGTGATGGATGTGGTCGCGGCC
GAGGTCTGGGGCCTGACTGCTCAGAACTTGCTGCTGTCTTGTTGCGGATGGATGGAAGGTTGGATGGATG
GGTGGATGGCCGTGGATGGCCGTGGATGCGCAGTGCCTTGCATACCCAAACCAGGTGGGAGCGTTTTGTT
GAGCATGACAGCCTGCAGCAGGAATATATGTGTGCCTATTTGTGTGGACAAAAATATTTACACTTAGGGT
TTGGAGCTATTCAAGAGGAAATGTCACAGAAGCAGCTAAACCAAGGACTGAGCACCCTCTGGATTCTGAA
TCTCAAGATGGGGGCAGGGCTGTGCTTGAAGGCCCTGCTGAGTCATCTGTTAGGGCCTTGGTTCAATAAA
GCACTGAGCAAGTTGAGAAA

>DEX0356_010.nt.3
GGCACGAGGAAGACAAAAAAAAAAATAAGCGTAACTTCCCTCAAAGCAACAACCCCCCCCCCCACCCCTCG
CCCGCTGCCTCCCTCGGCCCAGCCAGCTGTGCCGGCGTTTGTTGGCTGCCCTGCGCCCGGCCCTCCAGCC
AGCCTTCTGCCGGCCCCGCCGCGATGGAGGTGCCCCAGCCGGAGCCCGCGCCAGGCTCGGCTCTCAGTCC
AGCAGGCGTGTGCGGTGGCGCCCAGCGTCCGGGCCACCTCCCGGGCCTCCTGCTGGGATCTCATGGCCTC
CTGGGGTCCCCGGTGCGGGCGGCCGCTTCCTCGCCGGTCACCACCCTCACCCAGACCATGCACGACCTCG
CCGGGCTCGGCAGGTAGGACACCCCAAGGAGGCTGCATATGGGGGTGAGAGGCTAGGTCTGGAGTCCTGG
GCCTCCTGGAGAACTGGGCATGGTAGTCGAATCCAGGAGGTGGAGTCCGGGGAGGCATCCTGAGCAGTGG
AGAGGTACGCCCTTGTTCCCTCCTAGGGGCCCAGGACATGGGAAGTGAGACCTGGGCTGTGCCCCCCACA
CCTCCAGGTGACCTGCTCTCAGTCTGCACAAGCCTCTCCAAATCTTCTTTCCCCTCACTCCCTCCTCCCT
CATTGTGCCTCCACCACGCTGCCTCCGTAATCTCTGCCCGGATAGCCACCATTTTGCCCCAGGATGGGAA
TCTCAAGCCTGGAGAGGGAACTTAAGGGTGGCAGAGCCGTGGAGCACTCACACATCCATGCCCCTGCTCC
CTCCACCCCAGTTCTGCCCCCAGAATTAAGACAATTAATTAAACACTGATCCTGCTCCCTAAGTCTGCAA
GGGGAGAGCTGGGTTGTTCCAGTAGGAGAACCTGGTCTTGGTCCATCCTTCAGGAAGGGTCCTGGAAAGG
TGGGGGTCTGACAGTTTAGGGGAAGGCTCTCTGGGGCGGCTCTGCTAGACCACCAACCCTTAGGCTGGCC
TCCAGCCAGGCAGCCTTGCTCACACCCTGGGCTCACCTTGCCCTGCTCCTGAGAGGACCTTTCCTAAACC
AGTCAACGAAGAGAAGACGCTTTAGGGCCTCTGTAGAATGAACCTTCCTTGACCTGAGAAGAGGGGGCCC
TCAGGGCCATTGCTTTCCCGGTGTAGTGTTTCGCTCAGTTCTTTTCTCAGCAAGGGAGTGGGAACGTGGG
CTAGCCAGGCCTTGAGCCCCACGTTTACCTTTCTCCTTTCCCGGGTCCCTGTTCCCTTCCCAGCGAAACC
CCAAAGAGTCAGGTAGGGACCCTGCTCTTCCGCAGCCGCAGCCGCCTGACGCACCTATCCCTGTCTCGAC
GGGCATCCGAATCCTCCCTGTCGTCTGAATCCTCCGAATCTTCTGATGCAGGTCTCTGCATGGATTCCCC
CAGCCCTATGGACCCCCACATGGCGGAGCAGACGTTTGAACAGGCCATCCAGGCAGCCAGCCGGATCATT
CGAAACGAGCAGTTTGCCATCAGACGCTTCCAGTCTATGCCGGTGAGGCTGCTGGGCCACAGCCCCGTGC
TTCGGAACATCACCAACTCCCAGGCGCCCGACGGCCGGAGGAAGAGCGAGGCGGGCAGTGGAGCTGCCAG
CAGCTCTGGGGAAGACAAGGAGAATGTGCGCTTCTGGAAGGCCGGGGTGGGAGCTCTCCGGGAAGAGGAG
GGGGCATGCTGGGGTGGTTCCCTGGCATGTGAGGACCCTCCTCTCCCATCTTGGCTGCAGGATGGATTTG
TCTTCAAGATGCCATGGAAGCCCACACATCCCAGCTCCACCCATGCTCTGGCAGAGTGGGCCAGCCGCAG
GGAAGCCTTTGCCCAGAGACCCAGCTCGGCCCCCGACCTGATGTGTCTCAGTCCTGACCGGAAGATGGAA
GTGGAGGAGCTCAGCCCCCTGGCCCTAGGTCGCTTCTCTCTGACCCCTGCAGAGGGGGATACTGAGGAAG
ATGATGGATTTGTGGACATCCTAGAGAGTGACTTAAAGGTAAACAGCCTTGTCCCACCAGGCCCCTACCT
TCCTATCCCTGGGTTCGCCCAAAAGAAGAGAGCTCTGAGCCCTTGTGGCAGGACCATGATGTCATTCCAG
TGTCAGAAGAAGAATCTGAGGTAACTGAGTCACAGCCTAACCTCTGGCCAATGAGAGAAGAACAGGTGGC
TGGTGCCACAGTGGAGGGCGGTTTGGGAGGTAGGGGAGCTTCTCACCCCAACCTGGAGTTTGCCAGGGAA
ACAGGGAAGGCTGTGCTGGAGGATTCCGGGACTGGAGGGGTTTGGCCTATGCCTGTGGGTTGTGACCCTG
TCCTTGCTAATCTGGCCTCAGATGATGATGCAGTTCCCCCAGGCATGGAGAGTCTCATTAGTGCCCCACT
GGTCAAGACCTTGGAAAAGGAAGAGGAAAAGGACCTCGTCATGTACAGCAAGTGCCAGCGGCTCTTCCGC
TCTCCGTCCATGCCCTGCAGCGTGATCCGGCCCATCCTCAAGAGGCTGGAGCGGCCCCAGGACAGGGACA
CGCCCGTGCAGAATAAGCGGAGGCGGAGCGTGACCCCTCCTGAGGAGCAGCAGGAGGCTGAGGAACCTAA
AGCCCGCGTCCTCCGCTCAAAATCACTGTGTCACGATGAGATCGAGAACCTCCTGGACAGTGACCACCGA
GAGCTGATTGGAGATTACTCTAAGGCCTTCCTCCTACAGACAGTAGACGGAAAGCACCAAGACCTCAAGT
ACATCTCACCAGAAACGATGGTGGCCCTATTGACGGGCAAGTTCAGCAACATCGTGGATAAGTTTGTGAT
TGTAGACTGCAGATACCCCTATGAATATGAAGGCGGGCACATCAAGACTGCGGTGAACTTGCCCCTGGAA

CGCGACGCCGAGAGCTTCCTACTGAAGAGCCCCATCGCGCCCTGTAGCCTGGACAAGAGAGTCATCCTCA
TTTTCCACTGTGAATTCTCATCTGAGCGTGGGCCCCGCATGTGCCGTTTCATCAGGGAACGAGACCGTGC
TGTCAACGACTACCCCAGCCTCTACTACCCTGAGATGTATATCCTGAAAGGCGGCTACAAGGAGTTCTTC
CCTCAGCACCCGAACTTCTGTGAACCCCAGGACTACCGGCCCATGAACCACGAGGCCTTCAAGGATGAGC
TAAAGACCTTCCGCCTCAAGACTCGCAGCTGGGCTGGGGAGCGGAGCCGGCGGGAGCTCTGTAGCCGGCT
GCAGGACCAGTGAGGGGCCTGCGCCAGTCCTGCTACCTCCCTTGCCTTTCGAGGCCTGAAGCCAGCTGCC
CTATGGGCCTGCCGGGCTGAGGGCCTGCTGGAGGCCTCAGGTGCTGTCCATGGGAAAGATGGTGTGGGTG
TCCTGCCTGTCTGCCCCAGCCCAGATTCCCCTGTGTCATCCCATCATTTTCCATATCCTGGTGCCCCCCA
CCCCTGGAAGAGCCCAGTCTGTTGAGTTAGTTAAGTTGGGTTAATACCAGCTTAAAGGCAGTATTTTGTG
TCCTCCAGGAGCTTCTTGTTTCCTTGTTAGGGTTAACCCTTCATCTTCCTGTGTCCTGAAACGCTCCTTT
GTGTGTGTGTCAGCTGAGGCTGGGGGAGAGCCGTGGTCCCTGAGGATGGGTCAGAGCTAAACTCCTTCCT
GGCCTGAGAGTCAGCTCTCTGCCCTGTGTACTTCCCGGGCCAGGGCTGCCCCTAATCTCTGTAGGAACCG
TGGTATGTCTGCCATGTTGCCCCTTTCTCTTTTCCCCTTTCCTGTCCCACCATACGAGCACCTCCAGCCT
GAACAGAAGCTCTTACTCTTTCCTATTTCAGTGTTACCTGTGTGCTTGGTCTGTTTGACTTTACGCCCAT
CTCAGGACACTTCCGTAGACTGTTTAGGTTCCCCTGTCAAATATCAGTTACCCACTCGGTCCCAGTTTTG
TTGCCCCAGAAAGGGATGTTATTATCCTTGGGGGCTCCCAGGGCAAGGGTTAAGGCCTGAATCATGAGCC
TGCTGGAAGCCCAGCCCCTACTGCTGTGAACCCTGGGGCCTGACTGCTCAGAACTTGCTGCTGTCTTGTT
GCGGATGGATGGAAGGTTGGATGGATGGGTGGATGGCCGTGGATGGCCGTGGATGCGCAGTGCCTTGCAT
ACCCAAACCAGGTGGGAGCGTTTTGTTGAGCATGACAGCCTGCAGCAGGAATATATGTGTGCCTATTTGT
GTGGACAAAAATATTTACACTTAGGGTTTGGAGCTATTCAAGAGGAAATGTCACAGAAGCAGCTAAACCA
AGGACTGAGCACCCTCTGGATTCTGAATCTCAAGATGGGGGCAGGGCTGTGCTTGAAGGCCCTGCTGAGT
CATCTGTTAGGGCCTTGGTTCAATAAAGCACTGAGCAAGTTGAGAAA

>DEX0356_011.nt.1

TGCTTCCTGGGCGCCGTGGGCGCGGACTGCGCGGGCTGCGCGGGTGCCGAGGAGCGCGAGGCGCGGGGAA
GGCGCACCTGGGGTGGCCCTGGCGTGCGGGCGGCGACATGGAGGACGGCGTGCTCAAGGAGGGCTTCCTG
GTCAAGAGGGGCCACATTGTCCACAACTGGAAGGCGCGATGGTTCATCCTTCGGCAGAACACGCTGGTGT
ACTACAAGCTTGAGGGGGGTCGGAGAGTGACCCCTCCCAAGGGCCGGATCCTCCTGGATGGCTGCACCAT
CACCTGCCCCTGCCTGGAGTATGAAAACCGACCGCTCCTCATTAAGCTGAAGACTCAAACATCCACGGAG
TACTTCCTGGAGGCCTGTTCTCGAGAGGAGCGGGATGCCTGGGCCTTTGAGATCACCGGGGCTATTCATG
CAGGGCAGCCGGGGAAGGTCCAGCAGCTGCACAGCCTGAGAAACTCCTTCAAGCTGCCCCCGCACATCAG
CCTGCATCGCATTGTGGACAAGATGCACGATAGCAACACCGGAATCCGTTCAAGCCCCAACATGGAGCAG
GGAAGCACCTATAAAAAGACCTTCCTCGGCTCCTCCCTGGTGGACTGGCTCATCTCCAACAGCTTCACGG
CCAGCCGTCTGGAGGCGGTGACCCTGGCCTCCATGCTCATGGAGGAGAACTTCCTCAGGCCTGTGGGTGT
CCGAAGCATGGGAGCCATTCGCTCTGGGGATCTGGCCGAGCAGTTCCTGGATGACTCCACAGCCCTGTAC
ACTTTTGCTGAGAGCTACAAAAAGAAGATAAGCCCCAAGGAAGAAATTAGCCTGAGCACTGTGGAGTTAA
GTGGCACGGTGGTGAAACAAGGCTACCTGGCCAAGCAGGGACACAAGAGGAAAAACTGGAAGGTGCGTCG
CTTTGTTCTAAGGAAGGATCCAGCTTTCCTGCATTACTATGACCCTTCCAAAGAAGAGAACAGGCCAGTG
GGTGGGTTTTCTCTTCGTGGTTCACTCGTGTCTGCTCTGGAAGATAATGGCGTTCCCACTGGTAAGGTAC
AACTTCAGCCCTGGCCTGAAGATCTCAGGGGCCTCCCTCCTGCACCTCAGTAATGTCCCACTAATGAAAT
GACAGCTTTGGGAGCCGAGCAGGGCTTTCTCCATGAGACAGAGTGGTGTGAGTTTTGAAGTGAATTTCAG
AATTGAACAAAGGGTATGTTCTTGGCGTTTTAAGAATACAACTAAGTAATCCTTATGTTAATAATCTTTC
CCCTCAGTTTCTGTGCTGATGCCCCATAAGCTTCCGCCTAAAATAATAGGTGAGGCTTACACAATGCTCT
AGGGTTAACAACATACCGTAATATGGTATCATAATACTAATTTCACAATAGCCAGCCATAGTAAGAAGAT
ATTATTCCCATCTTGTGTCTGTGGAAACTGGGGGTTCCAATAAGTTAAACAATTTGAGATCAAACTCCAA
ATTCTATGCTTGCTTTTAACATAAACATACTGGATTCTTTGATGTAAATGAAATGGTAGATCTATGGCAT
GCTACCATCTTGAGCATAGGGAGCAGGCAAGAGGGCTTCCAGGAATATCCAGACATAACCATTGGAAGGC
CTTAGGGCACTGGTGAGGGACAGGGAATGGGAGGAAAAGGGGACCTCATCTGATAGGATTCTCTAACTGT
AAACCTTTTATTTAATAAATACATCATTGTATGTTTTTGTCCATGTCTCAGTTGACTTTCTGCTGGTTCA
AGATGAAAGCAGACATGGCTATTCAATCACTGATTTTTTTCCCCCTTTTTTCTGCCTAGGGGTTAAAGGG
AATGTCCAGGGAAACCTCTTCAAAGTGATTACTAAGGATGACACACACTATTACATTCAGGCCAGCAGCA
AGGCTGAGCGAGCCGAGTGGATTGAAGCTATCAAAAAGCTAACATGACAAGGACCTGAGGGAACCAGGAT
TCCTCCCTCCTACCAGATGACACAGACAAGAGTTCCTGGAGAATGGGAGTGTTAAGACTTTTGACTTCTT
TGTAAGTTTTGTACTGCTTTGGAGAGTGAATGCTGCCAAGAGTTCCTCAGATTACAAACAGCAGTGGTGC
CATTTCCTTCCCCATCTTCATGTTACAAACCTGGAAAGGCTAGAACAGCCATTAGGCGTCAGCATCTTGA
CTTTTCCCCAGCATCACAAACAGCCATTTCCTCGGGCACCAAAGTAGGTTCCCTTTGTTGGAACAATTAC
ACTGGCCATGCCATAATGTTGAATAAAACTCTCTTCTTATGAG

>DEX0356_012.nt.1

AGAGTCCGCCTATCAGCCAATGAAGAGACAGCAGTGAGAGCGGTTGCGCAGTGAAGGCTAGACCCGGTTT

ACTGGAATTGCTCTGGCGATCGAGGGGTCCTAGTACACCGCAATCATGTCTATTATGTCCTATAACGGAG
GGGCCGTCATGGCCATGAAGGGGAAGAACTGTGTGGCCATCGCTGCAGACAGGCGCTTCGGGATCCAGGC
CCAGATGGTGACCACGGACTTCCAGAAGATCTTTCCCATGGGTGACCGGCTGTACATCGGTCTGGCCGGG
CTCGCCACTGACGTCCAGACAGTTGCCCAGCGCCTCAAGTTCCGGCTGAACCTGTATGAGTTGAAGGAAG
GTCGGCAGATCAAACCTTATACCCTCATGAGCATGGTGGCCAACCTCTTGTATGAGAAACGGTTTGGCCC
TTACTACACTGAGCCAGTCATTGCCGGGTTGGACCCGAAGACCTTTAAGCCCTTCATTTGCTCTCTAGAC
CTCATCGGCTGCCCCATGGTGACTGATGACTTTGTGGTCAGTGGCACCTGCGCCGAACAAATGTACGGAA
TGTGTGAGTCCCTCTGGGAGCCCAACATGGTACGTTGGTGGCATGGAGAGGGGCTGGGCTCTGAGTTACC
CACCCTTGGTCATTAGGAAAAAAGTGTGTTTATGTGGCAGGTAATGGGGAGAATGGTGCAGGTGGGAGAG
AAAGCCACAGCTGCTGCTCCTTGGAAGGGTATCCAGAAATTCCATCTGCCAGGCTGGCTGGCTGGCTCCA
GAGGAGAGCATCTTCTGTTAGGAGACCACAGTTAGATATTTGCCACAGCAGTAAGAATTCCTTACATTGG
GTTTACAGGGTTCTCGCATCCATTCACTGAGCAAACATTTAACTCAGTGTAGGGTTAGAGGCAGCTCAAC
AGGTTGTCTGGGTTGGCAACCTGGATCCACCCACCACCTTAGTTATTGTTCTTGGAGAAGCCAGTTTCCC
TCTCTGCCTCCACTACCTTGACTGTAAAATGAGGATAAAAAAGTCTGCCTTAGAAGTGGTGCTGAG

>DEX0356_012.nt.2
AGAGTCCGCCTATCAGCCAATGAAGAGACAGCAGTGAGAGCGGTTGCGCAGTGAAGGCTAGACCCGGTTT
ACTGGAATTGCTCTGGCGATCGAGGGGTCCTAGTACACCGCAATCATGTCTATTATGTCCTATAACGGAG
GGGCCGTCATGGCCATGAAGGGGAAGAACTGTGTGGCCATCGCTGCAGACAGGCGCTTCGGGATCCAGGC
CCAGATGGTGACCACGGACTTCCAGAAGATCTTTCCCATGGGTGACCGGCTGTACATCGGTCTGGCCGGG
CTCGCCACTGACGTCCAGACAGTTGCCCAGCGCCTCAAGTTCCGGCTGAACCTGTATGAGTTGAAGGAAG
GTCGGCAGATCAAACCTTATACCCTCATGAGCATGGTGGCCAACCTCTTGTATGAGAAACGGTTTGGCCC
TTACTACACTGAGCCAGTCATTGCCGGGTTGGACCCGAAGACCTTTAAGCCCTTCATTTGCTCTCTAGAC
CTCATCGGCTGCCCCATGGTGACTGATGACTTTGTGGTCAGTGGCACCTGCGCCGAACAAATGTACCTCG
GCCGCGACCACGCTAAGCCGAATTCCAGCAC

>DEX0356_013.nt.1
CAAAACTTGACCGTACCCTCCAGTTCCCCCAGCCTGTGGGAAGCCCCAGCAGCTGAACCGGGTTGTGGGC
GGCGAGGACAGCACTGACAGCGAGTGGCCCTGGATCGTGAGCATCCAGAAGAATGGGACCCACCACTGCG
CAGGTTCTCTGCTCACCAGCCGCTGGGTGATCACTGCTGCCCACTGTTTCAAGGAGTATGTACAGCCGGC
CTGGGGCACTTGATCTTCTAAGGCCCTGGGCTTTGTGCCCCAGGCCCTGGGCTCCCTTTACAGGCTCTGC
TTCCGGGCCTCTGTTCAATCTTGGTGCCCCTCGATTCTACCTAGAAGCCTCTCCTGCCTTCAGGGGCCTT
CACACCCATGCTGTTCCCCCAGGTACCTCCTCCCAGACCAGGCCTGGCTGATCCTCCCCTTCCCTTTCCT
TCCAGCAACCTGAACAAACCATACCTGTTCTCTGTGCTGCTGGGGGCCTGGCAGCTGGGGAACCCTGGCT
CTCGGTCCCAGAAGGTGGGTGTTGCCTGGGTGGAGCCCCACCCTGTGTATTCCTGGAAGGAAGGTGCCTG
TGCAGACATTGCCCTGGTGCGTCTCGAGCGCTCCATACAGTTCTCAGAGCGGGTCCTGCCCATCTGCCTA
CCTGATGCCTCTATCCACCTCCCTCCAAACACCCACTGCTGGATCTCAGGCTGGGGGAGCATCCAAGATG
GAGTTCCCTTGCCCCACCCTCAGACCCTGCAGAAGCTGAAGGTTCCTATCATCGACTCGGAAGTCTGCAG
CCATCTGTACTGGCGGGGAGCAGGACAGGGACCCATCACTGAGGACATGCTGTGTGCCGGCTACTTGGAG
GGGGAGCGGGATGCTTGTCTGGGCGACTCCGGGGGCCCCCTCATGTGCCAGGTGGACGGCGCCTGGCTGC
TGGCCGGCATCATCAGCTGGGGCGAGGGCTGTGCCGAGCGCAACAGGCCCGGGGTCTACATCAGCCTCTC
TGCGCACCGCTCCTGGGTGGAGAAGATCGTGCAAGGGGTGCAGCTCCGCGGGCGCGCTCAGGGGGGTGGG
GCCCTCAGGGCACCGAGCCAGGGCTCTGGGGCCGCCGCGCGCTCCTAGGGCGCAGCGGGACGCGGGGCTC
GGATCTGAAAGGCGGCCAGATCCAGATCTGGATCTGGATCTGCGGCGGCCTCGGGCGGTTTCCCCCGCCG
TAAATAGGCTCATCTACTCTACCTCTGGGGGCCCGGACGGCTGCTGCGGAAAGGAAACCCCCTCCCCGAC
CCGCCCGACGGCCTCAGGCCCCGCCTCCAAGGCATCAGGCCCCGCCCAACGGCCTCATGTCCCCGCCCCC
ACGACTTCCGGCCCCGCCCCCGGGCCCCAGCGCTTTTGTGTATATAAATGTTAATGATTTTTATAGGTAT
TTGTAACCCTGCCCACATATCTTATTTATTCCTCCAATTTCAATAAATTATTTATTCTCCAGAAAAAATT
T

>DEX0356_014.nt.1
GCAAGTGTCGGTCGGGGCCTCGAGCTGCCTGAGCTGACACGAGGGGAGGGGTCTGTGTAGCCAACAGGTG
ACCGAAGGGCTTGCCTGCCCACAGCTTACTTGGCCAAGGGGTTTCTGAAGTTCCTGCCACAACCTGAGGT
TGTGGCTGGGAAGAGACGCAGAGGGCACAGAGCTGAGCCGCAGGCACAACTGGACAAAGCCAGAGCCCCA
GGCCCCTGTGGCCTGGGAAAGAGTAGCTCCCTCCAACCTCCCCCAGGGCCACCCGCTGCCCAAGTCCTTC
TCCTCCCCACCTTCTCCTTCGAACAAGAGGGAGGAGGAGGAGGAGGAGTTCAACTTCGAGGTCATCCCAC
CGCCGCCAGAGTTCAGCAATGACCCTGAGCCCCCGGCCCCGGCCCTCCAGTATCTGGGCCGCCAGAGCTC
CCCTCCCCGGAACAACTACTCAGACTTGAGGCAGCTCCCGAACGCTGGCCCCGGGGCGCCCCCGGCTCTC
GGCTTCTCGCGCTTTCCCGCGGGCGCGCGCTACGCCGGGGCTGGGGGCCTGGAGCGCTTCTCGGGAGGGG
GCCGCTCGCTCATAAAGAAGCGCCTGTACGTCGGGGAGCCGCACCGAGGCCCAGGGCTACCCCACGGTGG

CACCGGCCGCAGCCTGAGCTCTCCCAACTGCTTCGGGCCGCAGCCCGGAGGCCCCGAGATGCGGCGCGTG
AACTCGGCGGGTCGCGCGCCCCCCGGAGGCCTGCACGCGCCGAGGCTGTCCCTGGAGGGCGCCGCCCGGG
GCGCCGCGGAGGCCAAGCACAAAGCGCCCGGCAGCGCCGACTACGGCTTCGCCCCAGCTGCCGGCAGGTC
TCCCTACACCACCACCCGCTATGGAAGCCCCATCAACACGTTCACCGTGAGGCCTGGGACCCGCCATCCC
ATCTCCTATGTCTGCTCAGGGGCCCATCGGAAAGCCACCTCCTGAGCCAGGGGTTCTCTTAGCTCTACTC
CAACAAACTGTGTCTTTAAGGATCTGGGCACTTGTTTCTTGTGGGGGGTGGGAGGTACACAGCAGGTGTT
GGGCGACCTGACTCAGACACAGAGGAGTCTTTGATCCCCAAAGACCAACGATGAGAAGCAGTGGCCCAGG
AAGACCGAGAGTGGGCTTCTGTTTCCCAAAGCGCTGGTGTGACTGTGGACTGTTTACACGTGGGGGGTGG
GAGGGGAAGGAGGGTAGAGAGAAGGAACCATTAAGGACCTGGGCGCTGAGTTTCTGCTCTCTGGTACTAT
CTGCTTTAGCAAGAGTTATTTAAAACAGTTTTACAAGCACGGCCTACTGAGCAGGTTCTGAAAGCAAGGG
GTGGGGCACTGACTGAACTGGAACCCATCGAGATAGCTGTGATTGCCTGTCAGGCCCTTGCAAAGGGCTG
AGGCCCAGCTCTGTCTGGTGGGTGTGCAGGGCACCGAAGGGAGATCTTTAGCCTCCATCACCTGCCATCC
CAAGCAAGAGTGGACACCTGGGGTTCCTAGCCGCATTGAGGTGGGACCAGGAATGTGTATGTATAACATG
TGGAGAAGGACCATGTGTGGGGTGGGGAGTGAAGCTCCCACATGTGGAATGGAAGCCCTTTGCATTGGCC
AAGGGCAGAGGTGGGAGATTGATGGTTCATTTTTCCTTGGCCCTAGCCCTTAGCTACAGGGAGGCCTAGA
GAGAGGACACTTGGCAAGTTTAGGATGGGAAACAAGGTTAAGCCTGGCTTTGCTATTCAGACCTGTGCAG
CCCTGATTTTGGGGACTGGGCCCAAGGAGCCATGAGTCTTAGGCTTGACACTAGGGGCCTTACATGGCCA
GGTCTCATATGGGATTTGGAGAAGGCATTTTTCCTAACACTGATACTACAAGTCACCAATACTCTTGATG
CCTTAAGATGCTAGAAAGCCTTGGGAAAGTCAAGAGTGGTAGGATTGGTTGGCTGGGCGCGGTGGCTCACAC
TGCTGTTGGGGAATGTGTTTGGGAAAGTCAAGAGTGGTAGGATTGGTTGGCTGGGCGCGGTGGCTCACAC
CTGTAATCCCAGCACTTTGGGAGGCCGAGGCCGGTGGATCACTTGAGGTCAGAAGTTCGAGACCAGCCTG
CCCAAATGGTGAAACCCCATCTCTACTAAAAATACGGAAAAAAAAAAAAAAAAGAGTGCTAGGATTGGGCC
AGTTAAGAGCAGGGGAGACCTGGCTGGGCATGGTGGCTCACACCTGTAATTCTAGCACTTTGGGAGGCTG
AGGCGGGTGGATCATGAGGTCAAGAGATTAAGACCATCCTGGCCTACATGGCGAAACCCTGTCTCTACTA
AA

>DEX0356_015.nt.1
CGCCCCCGCCTCTGAGCTCCCTTCCCATGGCGGCCCTAGTGTTGGAGGACGGGTCGGTCCTGCGGGGCCA
GCCCTTTGGGGCCGCCGTGTCGACTGCCGGGGAAGTGGTGTTTCAAACCGGCATGGTCGGCTACCCCGAG
GCCCTCACTGATCCCTCCTACAAGGCACAGATCTTAGTGCTCACCTATCCTCTGATCGGCAACTATGGCA
TCCCCCCAGATGAAATGGATGAGTTCGGTCTCTGCAAGTGGTTTGAATCCTCGGGCATCCACGTAGCAGC
ACTGGTAGTGGGAGAGTGCTGTCCTACTCCCAGCCACTGGAGTGCCACCCGCACCCTGCATGAGTGGCTG
CAGCAGCATGGCATCCCTGGCTTGCAAGGAGTAGACACTCGGGAGCTGACCAAGAAGTTGCGGGAACAGG
GGTCTCTGCTGGGGAAGCTGGTCCAGAATGGAACAGAACCTTCATCCCTGCCATTCTTGGACCCCAATGC
CCGCCCCCTGGTACCAGAGGTCTCCATTAAGACTCCACGGGTATTCAATACAGGGGGTGCCCCTCGGATC
CTTGCTTTGGACTGTGGCCTCAAGTATAATCAGATCCGATGCCTCTGCCAGCGTGGGGCTGAGGTCACTG
TGGTACCCTGGGACCATGCACTAGACAGCCAAGAGTATGAGGGTCTCTTCTTAAGTAATGGGCCTGGTGA
CCCTGCCTCCTATCCCAGTGTCGTATCCACACTGAGCCGTGTTTTATCTGAGCCTAATCCCCGACCTGTC
TTTGGGATCTGCCTGGGACACCAGCTATTGGCCTTAGCCATTGGGGCCAAGACTTACAAGATGAGATATG
GGAACCGAGGCCATAACCAGCCCTGCTTGTTGGTGGGCTCTGGGCGCTGCTTTCTGACATCCCAGAACCA
TGGGTTTGCTGTGGAGACAGACTCACTGCCAGCAGACTGGGCTCCTCTCTTCACCAACGCCAATGATGGT
TCCAATGAAGGCATTGTGCACAACAGCTTGCCTTTCTTCAGTGTCCAGTTTCACCCAGAGCACCAAGCTG
GCCCTTCAGATATGGAACTGCTTTTCGATATCTTTCTGGAAACTGTGAAAGAGGCCACAGCTGGGAACCC
TGGGGGCCAGACAGTTAGAGAGCGGCTGACTGAGCGCCTCTGTCCCCCTGGGATTCCCACTCCCGGCTCT
GGACTTCCACCACCACGAAAGGTTCTGATCCTGGGCTCAGGGGGCCTCTCCATTGGCCAAGCTGGAGAAT
TTGACTACTCGGGCTCTCAGGCAATTAAGGCCCTGAAGGAGGAAAACATCCAGACGTTGCTGATCAACCC
CAATATTGCCACAGTGCAGACCTCCCAGGGGCTGGCCGACAAGGTCTATTTTCTTCCCATAACACCTCAT
TATGTAACCCAGGTGATACGTAATGAACGCCCCGATGGTGTGTTACTGACTTTTGGGGGCCAGACTGCTC
TGAACTGTGGTGTGGAGCTGACCAAGGCCGGGGTGCTGGCTCGGTATGGGGTCCGGGTCCTGGGCACAAC
AGTGGAGACCATTGAGCTGACCGAGGATCGACGGGCCTTTGCTGCCAGAATGGCAGAGATCGGAGAGCAT
GTGGCCCCGAGCGAGGCAGGAAATTCTCTTGAACAGGCCCAGGCAGCCGCTGAACGGCTGGGGTACCCTG
TGCTAGTGCGTGCAGCCTTTGCCGTGGGTGGCCTGGGCTCTGGCTTTGCCTCTAACAGGGAGGAGCTCTC
TGCTCTCGTGGCCCCAGCTTTTGCCCATACCAGCCAAGTGCTAGTAGACAAGTCTCTGAAGGGATGGAAG
GAGATTGAGTACGAGGTGGTGAGAGACGCCTATGGCAACTGTGTCACGGTGTGTAACATGGAGAACTTGG
ACCCACTGGGCATCCACACTGGTGAGTCCATAGTGGTGGCCCCTAGCCAGACACTGAATGACAGGGAGTA
TCAGCTCCTGAGGCAGACAGCTATCAAGGTGACCCAGCACCTGGGAATTGTTGGGGAGTGCAATGTGCAG
TATGCCTTGAACCCTGAGTCTGAGCAGTATTACATCATTGAAGTGAATGCCAGGCTCTCTCGCAGCTCTG
CCCTGGCCAGTAAGGCCACAGGTTATCCACTGGCTTATGTGGCAGCCAAGCTAGCATTGGGCATCCCTTT
GCCTGAGCTCAGGAACTCTGTGACAGGGGGTACAGCAGCCTTTGAACCCAGCGTGGATTATTGTGTGGTG
AAGATTCCTCGATGGGACCTTAGCAAGTTCCTGCGAGTCAGCACAAAGATTGGGAGCTGCATGAAGAGCG

TTGGTGAAGTCATGGGCATTGGGCGTTCATTTGAGGAGGCCTTCCAGAAGGCCCTGCGCATGGTGGATGA
GAACTGTGTGGGCTTTGATCACACAGTGAAACCAGTCAGCGATATGGAGTTGGAGACTCCAACAGATAAG
CGGATTTTTGTGGTGGCAGCTGCTTTGTGGGCTGGTTATTCAGTGGACCGCCTGTATGAGCTCACACGCA
TCGACCGCTGGTTCCTGCACCGAATGAAGCGTATCATCGCACATGCCCAGCTGCTAGAACAACACCGTGG
ACAGCCTTTGCCGCCAGACCTGCTGCAACAGGCCAAGTGTCTTGGCTTCTCAGACAAACAGATTGCCCTT
GCAGTTCTGAGCACAGAGCTGGCTGTTCGCAAGCTGCGTCAGGAACTGGGGATCTGTCCAGCAGTGAAAC
AGATTGACACAGTTGCAGCTGAGTGGCCAGCCCAGACAAATTACCTATACCTAACGTATTGGGGCACCAC
CCATGACCTCACCTTTCGAACACCTCATGTCCTAGTCCTTGGCTCTGGCGTCTACCGTATTGGCTCTAGC
GTTGAATTTGACTGGTGTGCTGTAGGCTGCATCCAGCAGCTCCGAAAGATGGGATATAAGACCATCATGG
TGAACTATAACCCAGAGACAGTCAGCACCGACTATGACATGTGTGATCGACTCTACTTTGATGAGATCTC
TTTTGAGGTGGTGATGGACATCTATGAGCTCGAGAACCCTGAAGGTGTGATCCTATCCATGGGTGGACAG
CTGCCCAACAACATGGCCATGGCGTTGCATCGGCAGCAGTGCCGGGTGCTGGGCACCTCCCCTGAAGCCA
TTGACTCGGCTGAGAACCGTTTCAAGTTTTCCCGGCTCCTTGACACCATTGGTATCAGCCAGCCTCAGTG
GAGGGAGCTCAGTGACCTCGAGTCTGCTCGCCAATTCTGCCAGACCGTGGGGTACCCCTGTGTGGTGCGC
CCCTCCTATGTGCTGAGCGGTGCTGCTATGAATGTGGCCTACACGGATGGAGACCTGGAGCGCTTCCTGA
GCAGCGCAGCAGCCGTCTCCAAAGAGCATCCCGTGGTCATCTCCAAGTTCATCCAGGAGGCTAAGGAGAT
TGACGTGGATGCCGTGGCCTCTGATGGTGTGGTGGCAGCCATCGCCATCTCTGAGCATGTGGAGAATGCA
GGTGTGCATTCAGGTGATGCGACGCTGGTGACCCCCCCACAAGATATCACTGCCAAAACCCTGGAGCGGA
TCAAAGCCATTGTGCATGCTGTGGGCCAGGAGCTACAGGTCACAGGACCCTTCAATCTGCAGCTCATTGC
CAAGGATGACCAGCTGAAAGTTATTGAATGCAACGTACGTGTCTCTCGCTCCTTCCCCTTCGTTTCCAAG
ACACTGGGTGTGGACCTAGTAGCCTTGGCCACGCGGGTCATCATGGGGGAAGAAGTGGAACCTGTGGGGC
TAATGACTGGTTCTGGAGTCGTGGGAGTAAAGGTGCCTCAGTTCTCCTTCTCCCGCTTGGCGGGTGCTGA
CGTGGTGTTGGGTGTGGAAATGACCAGTACTGGGGAGGTGGCCGGCTTTGGGGAGAGCCGCTGTGAGGCA
TACCTCAAGGCCATGCTAAGCACTGGCTTTAAGATCCCCAAGAAGAATATCCTGCTGACCATTGGCAGCT
ATAAGAACAAAAGCGAGCTGCTCCCAACTGTGCGGCTACTGGAGAGCCTGGGCTACAGCCTCTATGCCAG
TCTCGGCACAGCTGACTTCTACACTGAGCATGGCGTCAAGGTAACAGCTGTGGACTGGCACTTTGAGGAG
GCTGTGGATGGTGAGTGCCCACCACAGCGGAGCATCCTGGAGCAGCTAGCTGAGAAAAACTTTGAGCTGG
TGATTAACCTGTCAATGCGTGGAGCTGGGGGCCGGCGTCTCTCTTCCTTTGTCACCAAGGGCTACCGCAC
CCGACGCTTGGCCGCTGACTTCTCCGTGCCCCTAATCATCGATATCAAGTGCACCAAACTCTTTGTGGAG
GCCCTAGGCCAGATCGGGCCAGCCCCTCCTTTGAAGGTGCATGTTGACTGTATGACCTCCCAAAAGCTTG
TGCGACTGCCGGGATTGATTGATGTCCATGTGCACCTGCGGGAACCAGGTGGGACACATAAGGAGGACTT
TGCTTCAGGCACAGCCGCTGCCCTGGCTGGGGGTATCACCATGGTGTGTGCCATGCCTAATACCCGGCCC
CCCATCATTGACGCCCCTGCTCTGGCCCTGGCCCAGAAGCTGGCAGAGGCTGGCGCCCGGTGCGACTTTG
CGCTATTCCTTGGGGCCTCGTCTGAAAATGCAGGAACCTTGGGCACCGTGGCCGGGTCTGCAGCCGGGCT
GAAGCTTTACCTCAATGAGACCTTCTCTGAGCTGCGGCTGGACAGCGTGGTCCAGTGGATGGAGCATTTC
GAGACATGGCCCTCCCACCTCCCCATTGTGGCTCACGCAGAGCAGCAAACCGTGGCTGCTGTCCTCATGG
TGGCTCAGCTCACTCAGCGCTCAGTGCACATATGTCACGTGGCACGGAAGGAGGAGGTAAGAGTACACCT
GAGATCCTGCTGTCCCTGTTGCTTTCCCAGTAACACCAAAGGTCAGGGTAGTCCTTAGGGGCAGGAGACA
GCAGGAGGAGAGTCTGGAGACAGCAGGAGGAGGGTCTCGAGCCAGCACCCTTGCAGGTCTACATCATCAT
TTTTTTGAGCAGGGATGTTGGCCATTGGGACTTTGTGTAGGGCAAGGCATATGGGTGGTGCCTCTTCTGG
ATCTTCCCATTGTTCCCCAGATCCTGCTAATTAAAGCTGCAAAGGCACGGGGCTTGCCAGTGACCTGCGA
GGTGGCTCCCCACCACCTGTTCCTAAGCCATGATGACCTGGAGCGCCTGGGGCCTGGGAAGGGGGAGGTC
CGGCCTGAGCTTGGCTCCCGCCAGGATGTGGAAGCCCTGTGGGAGAACATGGCTGTCATCGACTGCTTTG
CCTCAGACCATGCTCCCCATACCTTGGAGGAGAAGTGTGGGTCCAGGCCCCCACCTGGGTTCCCAGGGTT
AGAGACCATGCTGCCACTACTCCTGACGGCTGTAAGCGAGGGCCGGCTCAGCCTGGACGACCTGCTGCAG
CGATTGCACCACAATCCTCGGCGCATCTTTCACCTGCCCCCGCAGGAGGACACCTATGTGGAGGTGGATC
TGGAGCATGAGTGGACAATTCCCAGCCACATGCCCTTCTCCAAGGCCCACTGGACACCTTTTGAAGGGCA
GAAAGTGAAGGGCACCGTCCGCCGTGTGGTCCTGCGAGGGGAGGTTGCCTATATCGATGGGCAGGTTCTG
GTACCCCCGGGCTATGGACAGGATGTACGGAAGTGGCCACAGGGGGCTGTTCCTCAGCTCCCACCCTCAG
CCCCTGCCACTAGTGAGATGACCACGACACCTGAAAGACCCCGCCGTGGCATCCCAGGGCTTCCTGATGG
CCGCTTCCATCTGCCGCCCCGAATCCATCGAGCCTCCGACCCAGGTTTGCCAGCTGAGGAGCCAAAGGAG
AAGTCCTCTCGGAAGGTAGCCGAGCCAGAGCTGATGGGAACCCCTGATGGCACCTGCTACCCTCCACCAC
CAGTACCGAGACAGGCATCTCCCCAGAACCTGGGGACCCCTGGCTTGCTGCACCCCCAGACCTCACCCCT
GCTGCACTCATTAGTGGGCCAACATATCCTGTCCGTCCAGCAGTTCACCAAGGATCAGATGTCTCACCTG
TTCAATGTGGCACACACACTGCGTATGATGGTGCAGAAGGAGCGGAGCCTCGACATCCTGAAGGGGAAGG
TCATGGCCTCCATGTTCTATGAAGTGAGCACACGGACCAGCAGCTCCTTTGCAGCAGCCATGGCCCGGCT
GGGAGGTGCTGTGCTCAGCTTCTCGGAAGCCACATCGTCCGTCCAGAAGGGCGAATCCCTGGCTGACTCC
GTGCAGACCATGAGCTGCTATGCCGACGTCGTCGTGCTCCGGCACCCCCAGCCTGGAGCAGTGGAGCTGG
CCGCCAAGCACTGCCGGAGGCCAGTGATCAATGCTGGGGATGGGGTCGGAGAGCACCCCACCCAGGCCCT
GCTGGACATCTTCACCATCCGTGAGGAGCTGGGAACTGTCAATGGCATGACGATCACGATGGTGGGTGAC

CTGAAGCACGGACGCACAGTACATTCCCTGGCCTGCCTGCTCACCCAGTATCGTGTCAGCCTGCGCTACG
TGGCACCTCCCAGCCTGCGCATGCCACCCACTGTGCGGGCCTTCGTGGCCTCCCGCGGCACCAAGCAGGA
GGAATTCGAGAGCATTGAGGAGGCGCTGCCTGACACTGATGTGCTCTACATGACTCGAATCCAGAAGGAA
CGATTTGGCTCTACCCAGGAGTACGAAGCTTGCTTTGGTCAGTTCATCCTCACTCCCCACATCATGACCC
GGGCCAAGAAGAAGATGGTGGTGATGCACCCGATGCCCCGTGTCAACGAGATAAGCGTGGAAGTGGACTC
GGATCCCCGCGCAGCCTACTTCCGCCAGGCTGAGAACGGCATGTACATCCGCATGGCTCTGTTAGCCACC
GTGCTGGGCCGTTTCTAGGGCCTGGCTTCCTCAGCCTCTTCTCTTTAGGCCCAGCTGCTGGGCAAGGAAT
TCCAGTGCCTCCTACGGGGGCAGCACACTTAGATATTCCTGGACATCCAGATAGCTCACATGTGCTGACC
ACACTTCAGGCTCTGGACTGGAGCTCTCTGGCATGGGGGTGGGGCCTCAGATGCTGGGGCCCAGTCTGCC
CCATCTTCATTCCTGCACCTTAAACCTGTACAGTCATTTTTCTACTGACTTAATAAACAGCCGAGCTGTC
CCTTGATGCTGAGTGTAGTAGAACAGAGCTTTCTTTTAGAAAATTGAGCAGATTCCCAAATTATAAGAGC
AGCCTCACCAGGCAGGGCTCTGGCTAGGGCTGCGTGCCCACACTCGGCATTTTCACACTCGTGACTCTTT
GGGACTCGGCAGGAGCGTGTTAACTCTGTGCTCTCTGAGTCCTGTGTCGTGCTGATGCCAGCGCTTACCT
GGGGGCTCTGGCTGGTGTCCAGCCCTCCTTCCTCAATTCAGTTGACTCAAAGACCTGACAGAGAGTTCTG
CCCTCTTCCTGACACTGGGAGAACACTCATCAGGGTAAAAGGAAGGCTGAGCACTGCGGCCTGTCTGAGG
CCAGGCCTGGACTCTTCCTCACTCTTTCTGGTTAGTCCTCCTCCAGTGGCTAAGGCTGCCATTGCATTCT
CAAAACCACCCTGGAGTTCTTCAGCCACTGTCACTATTAGGAGCCCCTCCTATTTCCCATCCCTGGGAGC
CTCTGCCTGACATTGGACAGAGGGGCAGATGCCATGGAGGCTGAGGAGCACTTTGGCTTGGAGACCCCAG
GGCAGCAGCCTAAGGGCAGTAGCTTATGTGTCTGTGGGTGAGGGCTTCTGAGGAGCAGCTGTGCCCTCCA
GTGAACTTGAGTAGCATTGGAGAAATCTCATTTCTTGTTCCAACAAAACTTAGCACTCTGTGGGTACAAA
GCACCATGTGCAGTGGGAGTACAAGGAAGAACAAGGCAATTCTCACCCTCAGTTTTCAGCCTCGGGTCAT
AAGGCTTCCAGTGGCTTCTACAGATTAAACTGCTGCACACTGGCACATGATAGTGCAAACTGCTCTGACA
CCATTCAGGAATCACCTACTTCATCCAGAGAGTAACAGCGTATTCTAGAGATGAGAAAAGGAGAAGGACC
TCGTTTTTTGAGAACTTAGTGTGAACAGGAACTTGATGCGTATTCTCATTTCTCACACCAGCCTTCCACC
CATTTCACAGATGTGGACACTGGCTCAGAGAAGTGGT

>DEX0356_015.nt.2
CGCCCCCGCCTCTGAGCTCCCTTCCCATGGCGGCCCTAGTGTTGGAGGACGGGTCGGTCCTGCGGGGCCA
GCCCTTTGGGGCCGCCGTGTCGACTGCCGGGGAAGTGGTGTTTCAAACCGGCATGGTCGGCTACCCCGAG
GCCCTCACTGATCCCTCCTACAAGGCACAGATCTTAGTGCTCACCTATCCTCTGATCGGCAACTATGGCA
TCCCCCCAGATGAAATGGATGAGTTCGGTCTCTGCAAGTGGTTTGAATCCTCGGGCATCCACGTAGCAGC
ACTGGTAGTGGGAGAGTGCTGTCCTACTCCCAGCCACTGGAGTGCCACCCGCACCCTGCATGAGTGGCTG
CAGCAGCATGGCATCCCTGGCTTGCAAGGAGTAGACACTCGGGAGCTGACCAAGAAGTTGCGGGAACAGG
GGTCTCTGCTGGGGAAGCTGGTCCAGAATGGAACAGAACCTTCATCCCTGCCATTCTTGGACCCCAATGC
CCGCCCCCCTGGTACCAGAGGTCTCCATTAAGACTCCACGGGTATTCAATACAGGGGGTGCCCCTCGGATC
CTTGCTTTGGACTGTGGCCTCAAGTATAATCAGATCCGATGCCTCTGCCAGCGTGGGGCTGAGGTCACTG
TGGTACCCTGGGACCATGCACTAGACAGCCAAGAGTATGAGGGTCTCTTCTTAAGTAATGGGCCTGGTGA
CCCTGCCTCCTATCCCAGTGTCGTATCCACACTGAGCCGTGTTTTATCTGAGCCTAATCCCCGACCTGTC
TTTGGGATCTGCCTGGGACACCAGCTATTGGCCTTAGCCATTGGGGCCAAGACTTACAAGATGAGATATG
GGAACCGAGGCCATAACCAGCCCTGCTTGTTGGTGGGCTCTGGGCGCTGCTTTCTGACATCCCAGAACCA
TGGGTTTGCTGTGGAGACAGACTCACTGCCAGCAGACTGGGCTCCTCTCTTCACCAACGCCAATGATGGT
TCCAATGAAGGCATTGTGCACAACAGCTTGCCTTTCTTCAGTGTCCAGTTTCACCCAGAGCACCAAGCTG
GCCCTTCAGATATGGAACTGCTTTTCGATATCTTTCTGGAAACTGTGAAAGAGGCCACAGCTGGGAACCC
TGGGGGCCAGACAGTTAGAGAGCGGCTGACTGAGCGCCTCTGTCCCCCTGGGATTCCCACTCCCGGCTCT
GGACTTCCACCACCACGAAAGGTTCTGATCCTGGGCTCAGGGGGCCTCTCCATTGGCCAAGCTGGAGAAT
TTGACTACTCGGGCTCTCAGGCAATTAAGGCCCTGAAGGAGGAAAACATCCAGACGTTGCTGATCAACCC
CAATATTGCCACAGTGCAGACCTCCCAGGGGCTGGCCGACAAGGTCTATTTTCTTCCCATAACACCTCAT
TATGTAACCCAGGTGATACGTAATGAACGCCCCGATGGTGTGTTACTGACTTTTGGGGGCCAGACTGCTC
TGAACTGTGGTGTGGAGCTGACCAAGGCCGGGGTGCTGGCTCGGTATGGGGTCCGGGTCCTGGGCACAAC
AGTGGAGACCATTGAGCTGACCGAGGATCGACGGGCCTTTGCTGCCAGAATGGCAGAGATCGGAGAGCAT
GTGGCCCCGAGCGAGGCAGGAAATTCTCTTGAACAGGCCCAGGCAGCCGCTGAACGGCTGGGGTACCCTG
TGCTAGTGCGTGCAGCCTTTGCCGTGGGTGGCCTGGGCTCTGGCTTTGCCTCTAACAGGGAGGAGCTCTC
TGCTCTCGTGGCCCCAGCTTTTGCCCATACCAGCCAAGTGCTAGTAGACAAGTCTCTGAAGGGATGGAAG
GAGATTGAGTACGAGGTGGTGAGAGACGCCTATGGCAACTGTGTCACGGTGTGTAACATGGAGAACTTGG
ACCCACTGGGCATCCACACTGGTGAGTCCATAGTGGTGGCCCCTAGCCAGACACTGAATGACAGGGAGTA
TCAGCTCCTGAGGCAGACAGCTATCAAGGTGACCCAGCACCTGGGAATTGTTGGGGAGTGCAATGTGCAG
TATGCCTTGAACCCTGAGTCTGAGCAGTATTACATCATTGAAGTGAATGCCAGGCTCTCTCGCAGCTCTG
CCCTGGCCAGTAAGGCCACAGGTTATCCACTGGCTTATGTGGCAGCCAAGCTAGCATTGGGCATCCCTTT
GCCTGAGCTCAGGAACTCTGTGACAGGGGGTACAGCAGCCTTTGAACCCAGCGTGGATTATTGTGTGGTG
AAGATTCCTCGATGGGACCTTAGCAAGTTCCTGCGAGTCAGCACAAAGATTGGGAGCTGCATGAAGAGCG

TTGGTGAAGTCATGGGCATTGGGCGTTCATTTGAGGAGGCCTTCCAGAAGGCCCTGCGCATGGTGGATGA
GAACTGTGTGGGCTTTGATCACACAGTGAAACCAGTCAGCGATATGGAGTTGGAGACTCCAACAGATAAG
CGGATTTTTGTGGTGGCAGCTGCTTTGTGGGCTGGTTATTCAGTGGACCGCCTGTATGAGCTCACACGCA
TCGACCGCTGGTTCCTGCACCGAATGAAGCGTATCATCGCACATGCCCAGCTGCTAGAACAACACCGTGG
ACAGCCTTTGCCGCCAGACCTGCTGCAACAGGCCAAGTGTCTTGGCTTCTCAGACAAACAGATTGCCCTT
GCAGTTCTGAGCACAGAGCTGGCTGTTCGCAAGCTGCGTCAGGAACTGGGGATCTGTCCAGCAGTGAAAC
AGATTGACACAGTTGCAGCTGAGTGGCCAGCCCAGACAAATTACCTATACCTAACGTATTGGGGCACCAC
CCATGACCTCACCTTTCGAACACCTCATGTCCTAGTCCTTGGCTCTGGCGTCTACCGTATTGGCTCTAGC
GTTGAATTTGACTGGTGTGCTGTAGGCTGCATCCAGCAGCTCCGAAAGATGGGATATAAGACCATCATGG
TGAACTATAACCCAGAGACAGTCAGCACCGACTATGACATGTGTGATCGACTCTACTTTGATGAGATCTC
TTTTGAGGTGGTGATGGACATCTATGAGCTCGAGAACCCTGAAGGTGTGATCCTATCCATGGGTGGACAG
CTGCCCAACAACATGGCCATGGCGTTGCATCGGCAGCAGTGCCGGGTGCTGGGCACCTCCCCTGAAGCCA
TTGACTCGGCTGAGAACCGTTTCAAGTTTTCCCGGCTCCTTGACACCATTGGTATCAGCCAGCCTCAGTG
GAGGGAGCTCAGTGACCTCGAGTCTGCTCGCCAATTCTGCCAGACCGTGGGGTACCCCTGTGTGGTGCGC
CCCTCCTATGTGCTGAGCGGTGCTGCTATGAATGTGGCCTACACGGATGGAGACCTGGAGCGCTTCCTGA
GCAGCGCAGCAGCCGTCTCCAAAGAGCATCCCGTGGTCATCTCCAAGTTCATCCAGGAGGCTAAGGAGAT
TGACGTGGATGCCGTGGCCTCTGATGGTGTGGTGGCAGCCATCGCCATCTCTGAGCATGTGGAGAATGCA
GGTGTGCATTCAGGTGATGCGACGCTGGTGACCCCCCCACAAGATATCACTGCCAAAACCCTGGAGCGGA
TCAAAGCCATTGTGCATGCTGTGGGCCAGGAGCTACAGGTCACAGGACCCTTCAATCTGCAGCTCATTGC
CAAGGATGACCAGCTGAAAGTTATTGAATGCAACGTACGTGTCTCTCGCTCCTTCCCCTTCGTTTCCAAG
ACACTGGGTGTGGACCTAGTAGCCTTGGCCACGCGGGTCATCATGGGGGGAAGAAGTGGAACCTGTGGGGC
TAATGACTGGTTCTGGAGTCGTGGGAGTAAAGGTGCCTCAGTTCTCCTTCTCCCGCTTGGCGGGTGCTGA
CGTGGTGTTGGGTGTGGAAATGACCAGTACTGGGGAGGTGGCCGGCTTTGGGGAGAGCCGCTGTGAGGCA
TACCTCAAGGCCATGCTAAGCACTGGCTTTAAGATCCCCAAGAAGAATATCCTGCTGACCATTGGCAGCT
ATAAGAACAAAAGCGAGCTGCTCCCAACTGTGCGGCTACTGGAGAGCCTGGGCTACAGCCTCTATGCCAG
TCTCGGCACAGCTGACTTCTACACTGAGCATGGCGTCAAGGTAACAGCTGTGGACTGGCACTTTGAGGAG
GCTGTGGATGGTGAGTGCCCACCACAGCGGAGCATCCTGGAGCAGCTAGCTGAGAAAAACTTTGAGCTGG
TGATTAACCTGTCAATGCGTGGAGCTGGGGGCCGGCGTCTCTCTTCCTTTGTCACCAAGGGCTACCGCAC
CCGACGCTTGGCCGCTGACTTCTCCGTGCCCCTAATCATCGATATCAAGTGCACCAAACTCTTTGTGGAG
GCCCTAGGCCAGATCGGGCCAGCCCCTCCTTTGAAGGTGCATGTTGACTGTATGACCTCCCAAAAGCTTG
TGCGACTGCCGGGATTGATTGATGTCCATGTGCACCTGCGGGAACCAGGTGGGACACATAAGGAGGACTT
TGCTTCAGGCACAGCCGCTGCCCTGGCTGGGGGTATCACCATGGTGTGTGCCATGCCTAATACCCGGCCC
CCCATCATTGACGCCCCTGCTCTGGCCCTGGCCCAGAAGCTGGCAGAGGCTGGCGCCCGGTGCGACTTTG
CGCTATTCCTTGGGGCCTCGTCTGAAAATGCAGGAACCTTGGGCACCGTGGCCGGGTCTGCAGCCGGGCT
GAAGCTTTACCTCAATGAGACCTTCTCTGAGCTGCGGCTGGACAGCGTGGTCCAGTGGATGGAGCATTTC
GAGACATGGCCCTCCCACCTCCCCATTGTGGCTCACGCAGAGCAGCAAACCGTGGCTGCTGTCCTCATGG
TGGCTCAGCTCACTCAGCGCTCAGTGCACATATGTCACGTGGCACGGAAGGAGGAGATCCTGCTAATTAA
AGCTGCAAAGGCACGGGGCTTGCCAGTGACCTGCGAGGTGGCTCCCCACCACCTGTTCCTAAGCCATGAT
GACCTGGAGCGCCTGGGGCCTGGGAAGGGGGAGGTCCGGCCTGAGCTTGGCTCCCGCCAGGATGTGGAAG
CCCTGTGGGAGAACATGGCTGTCATCGACTGCTTTGCCTCAGACCATGCTCCCCATACCTTGGAGGAGAA
GTGTGGGTCCAGGCCCCCACCTGGGTTCCCAGGGTTAGAGACCATGCTGCCACTACTCCTGACGGCTGTA
AGCGAGGGCCGGCTCAGCCTGGACGACCTGCTGCAGCGATTGCACCACAATCCTCGGCGCATCTTTCACC
TGCCCCCGCAGGAGGACACCTATGTGGAGGTGGATCTGGAGCATGAGTGGACAATTCCCAGCCACATGCC
CTTCTCCAAGGCCCACTGGACACCTTTTGAAGGGCAGAAAGTGAAGGGCACCGTCCGCCGTGTGGTCCTG
CGAGGGGAGGTTGCCTATATCGATGGGCAGGTTCTGGTACCCCCGGGCTATGGACAGGATGTACGGAAGT
GGCCACAGGGGGCTGTTCCTCAGCTCCCACCCTCAGCCCCTGCCACTAGTGAGATGACCACGACACCTGA
AAGACCCCGCCGTGGCATCCCAGGGCTTCCTGATGGCCGCTTCCATCTGCCGCCCCGAATCCATCGAGCC
TCCGACCCAGGTTTGCCAGCTGTGTTCCTCCGCCCAGGAGCTGGGATCCCACGGGGCAGCAGAGCGTGGG
CTGAGGAGCCAAAGGAGAAGTCCTCTCGGAAGGTAGCCGAGCCAGAGCTGATGGGAACCCCTGATGGCAC
CTGCTACCCTCCACCACCAGTACCGAGACAGGCATCTCCCCAGAACCTGGGGACCCCTGGCTTGCTGCAC
CCCCAGACCTCACCCCTGCTGCACTCATTAGTGGGCCAACATATCCTGTCCGTCCAGCAGTTCACCAAGG
ATCAGATGTCTCACCTGTTCAATGTGGCACACACACTGCGTATGATGGTGCAGAAGGAGCGGAGCCTCGA
CATCCTGAAGGGGAAGGTCATGGCCTCCATGTTCTATGAAGTGAGCACACGGACCAGCAGCTCCTTTGCA
GCAGCCATGGCCCGGCTGGGAGGTGCTGTGCTCAGCTTCTCGGAAGCCACATCGTCCGTCCAGAAGGGCG
AATCCCTGGCTGACTCCGTGCAGACCATGAGCTGCTATGCCGACGTCGTCGTGCTCCGGCACCCCCAGCC
TGGAGCAGTGGAGCTGGCCGCCAAGCACTGCCGGAGGCCAGTGATCAATGCTGGGGATGGGGTCGGAGAG
CACCCCACCCAGGCCCTGCTGGACATCTTCACCATCCGTGAGGAGCTGGGAACTGTCAATGGCATGACGA
TCACGATGGTGGGTGACCTGAAGCACGGACGCACAGTACATTCCCTGGCCTGCCTGCTCACCCAGTATCG
TGTCAGCCTGCGCTACGTGGCACCTCCCAGCCTGCGCATGCCACCCACTGTGCGGGCCTTCGTGGCCTCC
CGCGGCACCAAGCAGGAGGAATTCGAGAGCATTGAGGAGGCGCTGCCTGACACTGATGTGCTCTACATGA

CTCGAATCCAGAAGGAACGATTTGGCTCTACCCAGGAGTACGAAGCTTGCTTTGGTCAGTTCATCCTCAC
TCCCCACATCATGACCCGGGCCAAGAAGAAGATGGTGGTGATGCACCCGATGCCCCGTGTCAACGAGATA
AGCGTGGAAGTGGACTCGGATCCCCGCGCAGCCTACTTCCGCCAGGCTGAGAACGGCATGTACATCCGCA
TGGCTCTGTTAGCCACCGTGCTGGGCCGTTTCTAGGGCCTGGCTTCCTCAGCCTCTTCTCTTTAGGCCCA
GCTGCTGGGCAAGGAATTCCAGTGCCTCCTACGGGGGCAGCACACTTAGATATTCCTGGACATCCAGATA
GCTCACATGTGCTGACCACACTTCAGGCTCTGGACTGGAGCTCTCTGGCATGGGGGTGGGGCCTCAGATG
CTGGGGCCCAGTCTGCCCCATCTTCATTCCTGCACCTTAAACCTGTACAGTCATTTTTCTACTGACTTAA
TAAACAGCCGAGCTGTCCCTTGATGCTGAGTGTAGTAGAACAGAGCTTTCTTTTAGAAAATTGAGCAGAT
TCCCAAATTATAAGAGCAGCCTCACCAGGCAGGGCTCTGGCTAGGGCTGCGTGCCCACACTCGGCATTTT
CACACTCGTGACTCTTTGGGACTCGGCAGGAGCGTGTTAACTCTGTGCTCTCTGAGTCCTGTGTCGTGCT
GATGCCAGCGCTTACCTGGGGGCTCTGGCTGGTGTCCAGCCCTCCTTCCTCAATTCAGTTGACTCAAAGA
CCTGACAGAGAGTTCTGCCCTCTTCCTGACACTGGGAGAACACTCATCAGGGTAAAAGGAAGGCTGAGCA
CTGCGGCCTGTCTGAGGCCAGGCCTGGACTCTTCCTCACTCTTTCTGGTTAGTCCTCCTCCAGTGGCTAA
GGCTGCCATTGCATTCTCAAAACCACCCTGGAGTTCTTCAGCCACTGTCACTATTAGGAGCCCCTCCTAT
TTCCCATCCCTGGGAGCCTCTGCCTGACATTGGACAGAGGGGCAGATGCCATGGAGGCTGAGGAGCACTT
TGGCTTGGAGACCCCAGGGCAGCAGCCTAAGGGCAGTAGCTTATGTGTCTGTGGGTGAGGGCTTCTGAGG
AGCAGCTGTGCCCTCCAGTGAACTTGAGTAGCATTGGAGAAATCTCATTTCTTGTTCCAACAAAACTTAG
CACTCTGTGGGTACAAAGCACCATGTGCAGTGGGAGTACAAGGAAGAACAAGGCAATTCTCACCCTCAGT
TTTCAGCCTCGGGTCATAAGGCTTCCAGTGGCTTCTACAGATTAAACTGCTGCACACTGGCACATGATAG
TGCAAACTGCTCTGACACCATTCAGGAATCACCTACTTCATCCAGAGAGTAACAGCGTATTCTAGAGATG
AGAAAAGGAGAAGGACCTCGTTTTTTGAGAACTTAGTGTGAACAGGAACTTGATGCGTATTCTCATTTCT
CACACCAGCCTTCCACCCATTTCACAGATGTGGACACTGGCTCAGAGAAGTGGT

>DEX0356_015.nt.3
AACTCTGGCCAACGTTATCCCTCCAGACACCTGAAAGACCCCGCCGTGGCATCCCAGGGCTTCCTGATGG
CCGCTTCCATCTGCCGCCCCGAATCCATCGAGCCTCCGACCCAGGTTTGCCAGGTAAGAGTGGGGTTCCT
GTGACTCAGAGACTGTGTAGGGACAGGATCCACTTCTTCCCAGTGCCTCGCCTTTCTCTACTTACATGTC
CTCCTCTCCATCCCTTTATCCTCGTCTGATCTGCCGTGCCATCCTTTGCCTAAACAAGTCTCCCCGGTGT
GAGTAGACATCTCACAGCTTCTCACATGCCCTTTTTTGTTGTGGGCAGCTGTGTTCCTCCGCCCAGGAGC
TGGGATCCCACGGGGCAGCAGAGCGTGGGGTAAATCCAGGTTGTTGGTTGGTGTGAGTCTGGCCGTTCCT
CTTGCCTAGGATGCCTTTGCCAACTGGGAGAGCCCCGGGAGGGCACCACTGCTCCCATACAACACAGCCC
CATGGGAAGCCACCTGTCTGTCCCCAGAGCTGCTGCAGTCCCCTGCCCTCCCCTAACCTGCTATATTACT
GTGTTGTGAATTGGTTTAACATATACACTGTGATTCAGAGTTCCTGGGAATATGGGGTTTCTTTCCAAAC
CTCAGCCATAAATGTATATCTGTCCTCTTGTCCTGTTTGCAGCTGAGGAGCCAAAGGAGAAGTCCTCTCG
GAAGGTAGCCGAGCCAGAGCTGATGGGAACCCCTGATGGCACCTGCTACCCTCCACCACCAGTACCGAGA
CAGGCATCTCCCCAGAACCTGGGGACCCCTGGCTTGCTGCACCCCCAGACCTCACCCCTGCTGCACTCAT
TAGTGGGCCAACATATCCTGTCCGTCCAGCAGTTCACCAAGGATCAGATGTCTCACCTGTTCAATGTGGC
ACACACACTGCGTATGATGGTGCAGAAGGAGCGGAGCCTCGACATCCTGAAGGGGAAGGTCATGGCCTCC
ATGTTCTATGAAGTGAGCACACGGACCAGCAGCTCCTTTGCAGCAGCCATGGCCCGGCTGGGAGGTGCTG
TGCTCAGCTTCTCGGAAGCCACATCGTCCGTCCAGAAGGGCGAATCCCTGGCTGACTCCGTGCAGACCAT
GAGCTGCTATGCCGACGTCGTCGTGCTCCGGCACCCCCAGCCTGGAGCAGTGGAGCTGGCCGCCAAGCAC
TGCCGGAGGCCAGTGATCAATGCTGGGGATGGGGTCGGAGAGCACCCCACCCAGGCCCTGCTGGACATCT
TCACCATCCGTGAGGAGCTGGGAACTGTCAATGGCATGACGATCACGATGGTGGGTGACCTGAAGCACGG
ACGCACAGTACATTCCCTGGCCTGCCTGCTCACCCAGTATCGTGTCAGCCTGCGCTACGTGGCACCTCCC
AGCCTGCGCATGCCACCCACTGTGCGGGCCTTCGTGGCCTCCCGCGGCACCAAGCAGGAGGAATTCGAGA
GCATTGAGGAGGCGCTGCCTGACACTGATGTGCTCTACATGACTCGAATCCAGAAGGAACGATTTGGCTC
TACCCAGGAGTACGAAGCTTGCTTTGGTCAGTTCATCCTCACTCCCCACATCATGACCCGGGCCAAGAAG
AAGATGGTGGTGATGCACCCGATGCCCCGTGTCAACGAGATAAGCGTGGAAGTGGACTCGGATCCCCGCG
CAGCCTACTTCCGCCAGGCTGAGAACGGCATGTACATCCGCATGGCTCTGTTAGCCACCGTGCTGGGCCG
TTTCTAGGGCCTGGCTTCCTCAGCCTCTTCTCTTTAGGCCCAGCTGCTGGGCAAGGAATTCCAGTGCCTC
CTACGGGGGCAGCACACTTAGATATTCCTGGACATCCAGATAGCTCACATGTGCTGACCACACTTCAGGC
TCTGGACTGGAGCTCTCTGGCATGGGGGTGGGGCCTCAGATGCTGGGGCCCAGTCTGCCCCATCTTCATT
CCTGCACCTTAAACCTGTACAGTCATTTTTCTACTGACTTAATAAACAGCCGAGCTGTCCCTTGATGCTG
AGTGTAGTAGAACAGAGCTTTCTTTTAGAAAATTGAGCAGATTCCCAAATTATAAGAGCAGCCTCACCAG
GCAGGGCTCTGGCTAGGGCTGCGTGCCCACACTCGGCATTTTCACACTCGTGACTCTTTGGGACTCGGCA
GGAGCGTGTTAACTCTGTGCTCTCTGAGTCCTGTGTCGTGCTGATGCCAGCGCTTACCTGGGGGCTCTGG
CTGGTGTCCAGCCCTCCTTCCTCAATTCAGTTGACTCAAAGACCTGACAGAGAGTTCTGCCCTCTTCCTG
ACACTGGGAGAACACTCATCAGGGTAAAAGGAAGGCTGAGCACTGCGGCCTGTCTGAGGCCAGGCCTGGA
CTCTTCCTCACTCTTTCTGGTTAGTCCTCCTCCAGTGGCTAAGGCTGCCATTGCATTCTCAAAACCACCC
TGGAGTTCTTCAGCCACTGTCACTATTAGGAGCCCCTCCTATTTCCCATCCCTGGGAGCCTCTGCCTGAC

ATTGGACAGAGGGGCAGATGCCATGGAGGCTGAGGAGCACTTTGGCTTGGAGACCCCAGGGCAGCAGCCT
AAGGGCAGTAGCTTATGTGTCTGTGGGTGAGGGCTTCTGAGGAGCAGCTGTGCCCTCCAGTGAACTTGAG
TAGCATTGGAGAAATCTCATTTCTTGTTCCAACAAAACTTAGCACTCTGTGGGTACAAAGCACCATGTGC
AGTGGGAGTACAAGGAAGAACAAGGCAATTCTCACCCTCAGTTTTCAGCCTCGGGTCATAAGGCTTCCAG
TGGCTTCTACAGATTAAACTGCTGCACACTGGCACATGATAGTGCAAACTGCTCTGACACCATTCAGGAA
TCACCTACTTCATCCAGAGAGTAACAGCGTATTCTAGAGATGAGAAAAGGAGAAGGACCTCGTTTTTTGA
GAACTTAGTGTGAACAGGAACTTGATGCGTATTCTCATTTCTCACACCAGCCTTCCACCCATTTCACAGA
TGTGGACACTGGCTCAGAGAAGTGGT

>DEX0356_016.nt.1
AGCTGTTCCCCGCGCTGCGCGGAGCAGGGACCAGGCGGTTGCGGCGGCGACAGCCATGGCCGGCGCGCTG
GCAGGTCTGGCCGCGGGCTTGCAGGTCCCGCGGGTCGCGCCCAGCCCAGACTCGGACTCGGACACAGACT
CGGAGGACCCGAGTCTCCGGCGCAGCGCGGGCGGCTTGCTCCGCTCGCAGGTCATCCACAGCGGTCACTT
CATGGTGTCGTCGCCGCACAGCGACTCGCTGCCCCGGCGGCGCGACCAGGAGGGGTCCGTGGGGCCCTCC
GACTTCGGGCCGCGCAGTATCGACCCCACACTCACACGCCTCTTCGAGTGCTTGAGCCTGGCCTACAGTG
GCAAGCTGGTGTCTCCCAAGTGGAAGAATTTCAAAGGCCTCAAGCTGCTCTGCAGAGACAAGATCCGCCT
GAACAACGCCATCTGGAGGGCCTGGTATATCCAGTATGTGAAGCGGAGGAAGAGCCCCGTGTGTGGCTTC
GTGACCCCCCTGCAGGGGCCTGAGGCTGATGCGCACCGGAAGCCGGAGGCCGTGGTCCTGGAGGGGAACT
ACTGGAAGCGGCGCATCGAGGTGGTGATGCGGGAATACCACAAGTGGCGCATCTACTACAAGAAGCGGGT
CAGTGGGGGAGGGCCAGGGAGGCCCCAGAGCTTTCCTCCTGCGGCTGCCGGCTACCGCCCGCCTCGGAAG
ATCCCTGGAAAGGGGATCCTGACCCCCGAGCTTGCGCCCCTCGGGCCCTCCATTCAGTCCCGGGCCGACA
GCGCCACCGTGTGGCCACAGCGTCTCCTAGCGGCCTCCTTACCTAGGGGTCGGCTCCGTAAGCCCAGCAG
GGAAGATGACCTCCTGGCCCCTAAGCAGGCGGAAGGCAGGTGGCCGCCGCCGGAGCAATGGTGCAAACAG
CTCTTCTCCAGTGTGGTCCCCGTGCTGCTGGGGGACCCAGAGGAGGAGCCGGGTGGGCGGCAGCTCCTGG
ACCTCAATTGCTTTTTGTCCGACATCTCGGACACTCTCTTCACCATGACTCAGTCCGGCCCTTCGCCCCT
GCAGCTGCCGCCTGAGGATGCCTACGTCGGCAATGCTGACATGATCCAGCCGGACCTGACGCCACTGCAG
CCAAGCCTGGATGACTTCATGGACATCTCAGATTTCTTTACCAACTCCCGCCTCCCACAGCCGCCCATGC
CTTCAAACTTCCCAGAGCCCCCCAGCTTCAGCCCCGTGGTTGACTCCCTCTTCAGCAGTGGGACCCTGGG
CCCAGAGGTGCCCCCGGCTTCCTCGGCCATGACCCACCTCTCTGGACACAGCCGTCTGCAGGCTCGGAAC
AGCTGCCCTGGCCCCTTGGACTCCAGCGCCTTCCTGAGTTCTGATTTCCTCCTTCCTGAAGACCCCAAGC
CCCGGCTCCCACCCCCTCCTGTACCCCCACCTCTGCTGCATTACCCTCCCCCTGCCAAGGTGCCAGGCCT
GGAGCCCTGCCCCCCACCTCCCTTCCCTCCCATGGCACCACCCACTGCTTTGCTGCAGGAAGAGCCTCTC
TTCTCTCCCAGGTTTCCCTTCCCCACCGTCCCTCCTGCCCCAGGAGTGTCTCCGCTGCCTGCTCCTGCAG
CCTTCCCACCCACCCCACAGTCTGTCCCCAGCCCAGCCCCCACCCCCTTCCCCATAGAGCTTCTACCCTT
GGGGTATTCGGAGCCTGCCTTTGGGCCTTGCTTCTCCATGCCCAGAGGCAAGCCCCCCGCCCCATCCCCT
AGGGGACAGAAAGCCAGCCCCCCTACCTTAGCCCCTGCCACTGCCAGTCCCCCCACCACTGCGGGGAGCA
ACAACCCCTGCCTCACACAGCTGCTCACAGCAGCTAAGCCGGAGCAAGCCCTGGAGCCACCACTTGTATC
CAGCACCCTCCTCCGGTCCCCAGGGTCCCCGCAGGAGACAGTCCCTGAATTCCCCTGCACATTCCTTCCC
CCGACCCCGGCCCCTACACCGCCCCGGCCACCTCCAGGCCCGGCCACATTGGCCCCTTCCAGGCCCCTGC
TTGTCCCCAAAGCGGAGCGGCTCTCACCCCCAGCGCCCAGCGGTAAAGAGAGGGGTTGCAGGGATTGGGG
GAATCTAGGGGATGAGGGGAAGAGGAGGAAGAAGAAGGGGACTGGGTCTGGGGCCAGATGAGAGGAAGGG
GGCAATGGGGACCCTTTCCCCTGGGTCGGTCTTTCTGTGGGTCTGTCTGTCCCTACCACAGGCAGTGAAC
GGCGGCTGTCAGGGGACCTCAGCTCCATGCCAGGCCCTGGGACTCTGAGCGTCCGTGTCTCTCCCCCGCA
ACCCATCCTCAGCCGGGGCCGTCCAGACAGCAACAAGACCGAGAACCGGCGTATCACACACATCTCCGCG
GAGCAGAAGCGGCGCTTCAACATCAAGCTGGGGTTTGACACCCTTCATGGGCTCGTGAGCACACTCAGTG
CCCAGCCCAGCCTCAAGGTGAGCAAAGCTACCACGCTGCAGAAGACAGCTGAGTACATCCTTATGCTACA
GCAGGAGCGTGCGGGCTTGCAGGAGGAGGCCCAGCAGCTGCGGGATGAGATTGAGGAGCTCAATGCCGCC
ATTAACCTGTGCCAGCAGCAGCTGCCCGCCACAGGGGTACCCATCACACACCAGCGTTTTGACCAGATGC
GAGACATGTTTGATGACTACGTCCGAACCCGTACGCTGCACAACTGGAAGTTCTGGGTGTTCAGCATCCT
CATCCGGCCTCTGTTTGAGTCCTTCAACGGGATGGTGTCCACGGCAAGTGTGCACACCCTCCGCCAGACC
TCACTGGCCTGGCTGGACCAGTACTGCTCTCTGCCCGCTCTCCGGCCAACTGTCCTGAACTCCCTACGCC
AGCTGGGCACATCTACCAGTATCCTGACCGACCCGGGCCGCATCCCTGAGCAAGCCACACGGGCAGTCAC
AGAGGGCACCCTTGGCAAACCTTTATAGTCCTGGCCAGACCCTGCTGCTCACTCAGCTGCCCTGGGGGCT
GCTTTCCCTGGGCACGGGCTCCAGGGATCATCTCTGGGCACTCCCTTCCTGCCCCAGGCCCTGGCTCTGC
CCTTCCCTGGGGGGTGGAGCAGGGTCCAGGTTTCACACTTGCCACCTCCTGGAGGTCAAGAAGAGCAGAG
TCCCCGTCCCTGCTCTGCCACTGTGCTCCAGCACCGTGACCTTGGGTGACTCGTCCGCTGTCTTTGGACC
GCTGTGTTTCAATCTGCAAAATGGGGATGGGGAAGGTTCAATCAGCAGATGACCCCCAGGCCTTGGCAGC
TGTGACATTGGGGGCCTAGGCTGGCAACTCCGGGGGCTCAACGGTGGAAAGAGGAGGATGCTGTTTCTCT
GTCACCTCCACTTGCTCCCCGACAGGTGGGGCACAGACCTCTGTTCCTGAGCAGAGAAGCAGAAAAGGAG
GTTCCCTCTCTCTGCTCCTTCACTGCTGACCCAGAGGGGCTGCAGGATGGTTTCCCCTGGGAGAGGCCAG

GAGGGCCTGATCCCAGGAGACACCAGGGCCAGAGTGACCACAGCAGGGCAGGCATCATGTGTGTGTGTGT
GTGTGGATGTGTGTGTGTGGGTTTTGTAAAGAATTCTTGACCAATAAAAGCAAAAACTGTCTGCTGG

>DEX0356_016.nt.2
CTCATCGAAGCCCCTGGGGACAGAAAGAGGGTCCCTGGGATAAAGCTTAGTCTTAGTGACCTCCTAATTC
CTTCCCCACGGGGTCCTGGGCCTGTCAATCAAGAGGTGGAGGCCTATGGGGCGGGAACTAGGGAATGAGG
AACCACCCCCACCCCTGTTCCAGTCCCGGGAGGGGGTGGGGGCTGCCCAGGGGCAGAAGTGTGCATGGCA
GTGCCCTGGGAGCAGGCCAGCCTGGGAGGGCAGCGGCCCTGTCTGGGCACACAGCGCCCGGCCACAGATT
TCTTTACCAACTCCCGCCTCCCACAGCCGCCCATGCCTTCAAACTTCCCAGAGCCCCCCAGCTTCAGCCC
CGTGGTTGACTCCCTCTTCAGCAGTGGGACCCTGGGCCCAGAGGTGCCCCCGGCTTCCTCGGCCATGACC
CACCTCTCTGGACACAGCCGTCTGCAGGCTCGGAACAGCTGCCCTGGCCCCTTGGACTCCAGCGCCTTCC
TGAGTTCTGATTTCCTCCTTCCTGAAGACCCCAAGCCCCGGCTCCCACCCCCTCCTGTACCCCCACCTCT
GCTGCATTACCCTCCCCCTGCCAAGGTGCCAGGCCTGGAGCCCTGCCCCCCACCTCCCTTCCCTCCCATG
GCACCACCCACTGCTTTGCTGCAGGAAGAGCCTCTCTTCTCTCCCAGGTTTCCCTTCCCCACCGTCCCTC
CTGCCCCAGGAGTGTCTCCGCTGCCTGCTCCTGCAGCCTTCCCACCCACCCCACAGTCTGTCCCCAGCCC
AGCCCCCACCCCCTTCCCCATAGAGCTTCTACCCTTGGGGTATTCGGAGCCTGCCTTTGGGCCTTGCTTC
TCCATGCCCAGAGGCAAGCCCCCCGCCCCATCCCCTAGGGGACAGAAAGCCAGCCCCCCTACCTTAGCCC
CTGCCACTGCCAGTCCCCCCACCACTGCGGGGAGCAACAACCCCTGCCTCACACAGCTGCTCACAGCAGC
TAAGCCGGAGCAAGCCCTGGAGCCACCACTTGTATCCAGCACCCTCCTCCGGTCCCCAGGGTCCCCGCAG
GAGACAGTCCCTGAATTCCCCTGCACATTCCTTCCCCCGACCCCGGCCCCTACACCGCCCCGGCCACCTC
CAGGCCCGGCCACATTGGCCCCTTCCAGGCCCCTGCTTGTCCCCAAAGCGGAGCGGCTCTCACCCCCAGC
GCCCAGCGGTAAAGAGAGGGGTTGCAGGGATTGGGGGAATCTAGGGGATGAGGGGAAGAGGAGGAAGAAG
AAGGGGACTGGGTCTGGGGCCAGATGAGAGGAAGGGGGCAATGGGGACCCTTTCCCCTGGGTCGGTCTTT
CTGTGGGTCTGTCTGTCCCTACCACAGGCAGTGAACGGCGGCTGTCAGGGGACCTCAGCTCCATGCCAGG
CCCTGGGACTCTGAGCGTCCGTGTCTCTCCCCCGCAACCCATCCTCAGCCGGGGCCGTCCAGACAGCAAC
AAGACCGAGAACCGGCGTATCACACACATCTCCGCGGAGCAGAAGCGGCGCTTCAACATCAAGCTGGGGT
TTGACACCCTTCATGGGCTCGTGAGCACACTCAGTGCCCAGCCCAGCCTCAAGGTGAGCAAAGCTACCAC
GCTGCAGAAGACAGCTGAGTACATCCTTATGCTACAGCAGGAGCGTGCGGGCTTGCAGGAGGAGGCCCAG
CAGCTGCGGGATGAGATTGAGGAGCTCAATGCCGCCATTAACCTGTGCCAGCAGCAGCTGCCCGCCACAG
GGGTACCCATCACACACCAGCGTTTTGACCAGATGCGAGACATGTTTGATGACTACGTCCGAACCCGTAC
GCTGCACAACTGGAAGTTCTGGGTGTTCAGCATCCTCATCCGGCCTCTGTTTGAGTCCTTCAACGGGATG
GTGTCCACGGCAAGTGTGCACACCCTCCGCCAGACCTCACTGGCCTGGCTGGACCAGTACTGCTCTCTGC
CCGCTCTCCGGCCAACTGTCCTGAACTCCCTACGCCAGCTGGGCACATCTACCAGTATCCTGACCGACCC
GGGCCGCATCCCTGAGCAAGCCACACGGGCAGTCACAGAGGGCACCCTTGGCAAACCTTTATAGTCCTGG
CCAGACCCTGCTGCTCACTCAGCTGCCCTGGGGGCTGCTTTCCCTGGGCACGGGCTCCAGGGATCATCTC
TGGGCACTCCCTTCCTGCCCCAGGCCCTGGCTCTGCCCTTCCCTGGGGGGTGGAGCAGGGTCCAGGTTTC
ACACTTGCCACCTCCTGGAGGTCAAGAAGAGCAGAGTCCCCGTCCCTGCTCTGCCACTGTGCTCCAGCAC
CGTGACCTTGGGTGACTCGTCCGCTGTCTTTGGACCGCTGTGTTTCAATCTGCAAAATGGGGATGGGGAA
GGTTCAATCAGCAGATGACCCCCAGGCCTTGGCAGCTGTGACATTGGGGGCCTAGGCTGGCAACTCCGGG
GGCTCAACGGTGGAAAGAGGAGGATGCTGTTTCTCTGTCACCTCCACTTGCTCCCCGACAGGTGGGGCAC
AGACCTCTGTTCCTGAGCAGAGAAGCAGAAAAGGAGGTTCCCTCTCTCTGCTCCTTCACTGCTGACCCAG
AGGGGCTGCAGGATGGTTTCCCCTGGGAGAGGCCAGGAGGGCCTGATCCCAGGAGACACCAGGGCCAGAG
TGACCACAGCAGGGCAGGCATCATGTGTGTGTGTGTGTGGATGTGTGTGTGTGGGTTTTGTAAAGAAT
TCTTGACCAATAAAAGCAAAAACTGTCTGCTGG

>DEX0356_016.nt.3
AGCTGTTCCCCGCGCTGCGCGGAGCAGGGACCAGGCGGTTGCGGCGGCGACAGCCATGGCCGGCGCGCTG
GCAGGTCTGGCCGCGGGCTTGCAGGTCCCGCGGGTCGCGCCCAGCCCAGACTCGGACTCGGACACAGACT
CGGAGGACCCGAGTCTCCGGCGCAGCGCGGGCGGCTTGCTCCGCTCGCAGGTCATCCACAGCGGTCACTT
CATGGTGTCGTCGCCGCACAGCGACTCGCTGCCCCGGCGGCGCGACCAGGAGGGGTCCGTGGGGCCCTCC
GACTTCGGGCCGCGCAGTATCGACCCCACACTCACACGCCTCTTCGAGTGCTTGAGCCTGGCCTACAGTG
GCAAGCTGGTGTCTCCCAAGTGGAAGAATTTCAAAGGCCTCAAGCTGCTCTGCAGAGACAAGATCCGCCT
GAACAACGCCATCTGGAGGGCCTGGTATATCCAGTATGTGAAGCGGAGGAAGAGCCCCGTGTGTGGCTTC
GTGACCCCCCTGCAGGGGCCTGAGGCTGATGCGCACCGGAAGCCGGAGGCCGTGGTCCTGGAGGGGAACT
ACTGGAAGCGGCGCATCGAGGTGGTGATGCGGGAATACCACAAGTGGCGCATCTACTACAAGAAGCGGGT
CAGTGGGGGAGGGCCAGGGAGGCCCCAGAGCTTTCCTCCTGCGGCTGCCGGCTACCGCCCGCCTCGGAAG
ATCCCTGGAAAGGGGATCCTGACCCCCGAGCTTGCGCCCCTCGGGCCCTCCATTCAGTCCCGGGCCGACA
GCGCCACCGTGTGGCCACAGCGTCTCCTAGCGGCCTCCTTACCTAGGGGTCGGGTGAGCTCCTGATGGGA
AATGGGGGATCTCATCGCTTGTGAGTAGAGGAGACTTTGGGGGGAAAGTGATGGAGGATGGGGCAAGGGA
TCCGGTGTCCAACTCTGTGTGTCCCTGCAGCTCCGTAAGCCCAGCAGGGAAGATGACCTCCTGGCCCCTA

AGCAGGCGGAAGGCAGGTGGCCGCCGCCGGAGCAATGGTGCAAACAGCTCTTCTCCAGTGTGGTCCCCGT
GCTGCTGGGGGACCCAGAGGAGGAGCCGGGTGGGCGGCAGCTCCTGGACCTCAATTGCTTTTTGTCCGAC
ATCTCGGACACTCTCTTCACCATGACTCAGTCCGGCCCTTCGCCCCTGCAGCTGCCGCCTGAGGATGCCT
ACGTCGGCAATGCTGACATGATCCAGCCGGACCTGACGCCACTGCAGCCAAGCCTGGATGACTTCATGGA
CATCTCAGATTTCTTTACCAACTCCCGCCTCCCACAGCCGCCCATGCCTTCAAACTTCCCAGAGCCCCCC
AGCTTCAGCCCCGTGGTTGACTCCCTCTTCAGCAGTGGGACCCTGGGCCCAGAGGTGCCCCCGGCTTCCT
CGGCCATGACCCACCTCTCTGGACACAGCCGTCTGCAGGCTCGGAACAGCTGCCCTGGCCCCTTGGACTC
CAGCGCCTTCCTGAGTTCTGATTTCCTCCTTCCTGAAGACCCCAAGCCCCGGCTCCCACCCCCTCCTGTA
CCCCCACCTCTGCTGCATTACCCTCCCCCTGCCAAGGTGCCAGGCCTGGAGCCCTGCCCCCCACCTCCCT
TCCCTCCCATGGCACCACCCACTGCTTTGCTGCAGGAAGAGCCTCTCTTCTCTCCCAGGTTTCCCTTCCC
CACCGTCCCTCCTGCCCCAGGAGTGTCTCCGCTGCCTGCTCCTGCAGCCTTCCCACCCACCCCACAGTCT
GTCCCCAGCCCAGCCCCCACCCCCTTCCCCATAGAGCTTCTACCCTTGGGGTATTCGGAGCCTGCCTTTG
GGCCTTGCTTCTCCATGCCCAGAGGCAAGCCCCCCGCCCCATCCCCTAGGGGACAGAAAGCCAGCCCCCC
TACCTTAGCCCCTGCCACTGCCAGTCCCCCCACCACTGCGGGGAGCAACAACCCCTGCCTCACACAGCTG
CTCACAGCAGCTAAGCCGGAGCAAGCCCTGGAGCCACCACTTGTATCCAGCACCCTCCTCCGGTCCCCAG
GGTCCCCGCAGGAGACAGTCCCTGAATTCCCCTGCACATTCCTTCCCCCGACCCCGGCCCCTACACCGCC
CCGGCCACCTCCAGGCCCGGCCACATTGGCCCCTTCCAGGCCCCTGCTTGTCCCCAAAGCGGAGCGGCTC
TCACCCCCAGCGCCCAGCGGTAAAGAGAGGGGTTGCAGGGATTGGGGGAATCTAGGGGATGAGGGGAAGA
GGAGGAAGAAGAAGGGGACTGGGTCTGGGGCCAGATGAGAGGAAGGGGGCAATGGGGACCCTTTCCCCTG
GGTCGGTCTTTCTGTGGGTCTGTCTGTCCCTACCACAGGCAGTGAACGGCGGCTGTCAGGGGACCTCAGC
TCCATGCCAGGCCCTGGGACTCTGAGCGTCCGTGTCTCTCCCCCGCAACCCATCCTCAGCCGGGGCCGTC
CAGACAGCAACAAGACCGAGAACCGGCGTATCACACACATCTCCGCGGAGCAGAAGCGGCGCTTCAACAT
CAAGCTGGGGTTTGACACCCTTCATGGGCTCGTGAGCACACTCAGTGCCCAGCCCAGCCTCAAGGTGAGC
AAAGCTACCACGCTGCAGAAGACAGCTGAGTACATCCTTATGCTACAGCAGGAGCGTGCGGGCTTGCAGG
AGGAGGCCCAGCAGCTGCGGGATGAGATTGAGGAGCTCAATGCCGCCATTAACCTGTGCCAGCAGCAGCT
GCCCGCCACAGGGGTACCCATCACACACCAGCGTTTTGACCAGATGCGAGACATGTTTGATGACTACGTC
CGAACCCGTACGCTGCACAACTGGAAGTTCTGGGTGTTCAGCATCCTCATCCGGCCTCTGTTTGAGTCCT
TCAACGGGATGGTGTCCACGGCAAGTGTGCACACCCTCCGCCAGACCTCACTGGCCTGGCTGGACCAGTA
CTGCTCTCTGCCCGCTCTCCGGCCAACTGTCCTGAACTCCCTACGCCAGCTGGGCACATCTACCAGTATC
CTGACCGACCCGGGCCGCATCCCTGAGCAAGCCACACGGGCAGTCACAGAGGGCACCCTTGGCAAACCTT
TATAGTCCTGGCCAGACCCTGCTGCTCACTCAGCTGCCCTGGGGGCTGCTTTCCCTGGGCACGGGCTCCA
GGGATCATCTCTGGGCACTCCCTTCCTGCCCCAGGCCCTGGCTCTGCCCTTCCCTGGGGGGTGGAGCAGG
GTCCAGGTTTCACACTTGCCACCTCCTGGAGGTCAAGAAGAGCAGAGTCCCCGTCCCTGCTCTGCCACTG
TGCTCCAGCACCGTGACCTTGGGTGACTCGTCCGCTGTCTTTGGACCGCTGTGTTTCAATCTGCAAAATG
GGGATGGGGAAGGTTCAATCAGCAGATGACCCCCAGGCCTTGGCAGCTGTGACATTGGGGGCCTAGGCTG
GCAACTCCGGGGGCTCAACGGTGGAAAGAGGAGGATGCTGTTTCTCTGTCACCTCCACTTGCTCCCCGAC
AGGTGGGGCACAGACCTCTGTTCCTGAGCAGAGAAGCAGAAAAGGAGGTTCCCTCTCTCTGCTCCTTCAC
TGCTGACCCAGAGGGGCTGCAGGATGGTTTCCCCTGGGAGAGGCCAGGAGGGCCTGATCCCAGGAGACAC
CAGGGCCAGAGTGACCACAGCAGGGCAGGCATCATGTGTGTGTGTGTGTGTGGATGTGTGTGTGTGGGTT
TTGTAAAGAATTCTTGACCAATAAAAGCAAAAACTGTCTGCTGGAAAAAAAAAAAATACTTGAAAAACTG
AAAGCTTAATGGAGCGTTACTAACCTTAAAAGAAAGGAGTACTTTGACATTTAAAAA

>DEX0356_017.nt.1
GTTCTCCCTTGTCCTGGCCCTGGGCCTCTAGTCACTTCCTGCTTCCCACCAACAGTTTCTCCTTACCCCC
ACCCAGGTCCTGGATGGTCGCCGGCCCACAGGGGGGCGACTGGAGGTAATGGTCCGGATTCGGGAGCCAC
TGACAGCCCAGCAGTTGGAGACGACGACAGAGAGCAGGTTGCTGGGCCCAAAGGGAAGGCCCCTCCTGTG
CCTGCCCCTGCAAGGGAGTCAGGGAACAGATCAGCCCGGCCCCTGCATAGCCTCAGTGTGCTGGCGTTTG
ACCAAGAGCGTCTGGAGCGGAAGATCCTGGCCCTCAGGCAGGCGCGGCGGCCGGTGCCCCCAGAAGTGGC
CCAGCAGTACCAGGACATCATGCAACGCAGCCAGTGGCAGAGGGCACAGCTGGAGCAGGGGGGTGTGGGC
ATCCGACGGGAATACGCAGCCCAGCTGGAGCGGCAGCTGCAGTTCTACACGGAGGCTGCCCGGCGCCTGG
GCAACGATGGCAGCAGGGATGCTGCAAAGGAGGCGCTCTATAGGCGGAATCTGGTAGAGAGTGAGCTGCA
GCGGCTCCGCAGGTGAGGAGCCCATGGGGCGGGCAGCCCCCAGAAAGCGGGCAGCAGGCCCCGATACCGG
GAAGAGCCGACACAGCCACGAACCAGACAAGCAGACAATCAGCGGACAATCGGTTCTGGACTCACCCCTC
ATCCGGGCCCCCAGCCCCGCCAGAGCCTCCGTGGCTGCGGGTGTTGGGAACCATGCCTGCCAGCCAGTAT
GTGCCCCTCACCCAGGCCTGGCTGGGCCCTGGAGAGTCCTGTTTGCACAGCCCAGGGGTGTCCGGCCTCT
GGCCCGCCCCGGAGCAGGGAGGGTGGCTGGGGCCAAGCCCCGAGGGCCCCTGCAAGCACTTTACTTCCTG
TTCCTCCCCAGCCTTAACCCCAAAGCCCTCCTGCACCCCAAAGAAGCCACTGAGGCTGGCCGAGCCACAC
TGTCTCCCCAGGGGCGTCGACCTGGCCCAGCTGGGTCCCCAGGGCCAGCACATGGAATAAAATAGCCAGG
GCCACACTCA

>DEX0356_017.nt.2
GGGAGGGGCTCGGGGATGGGCACCTGGAGGGGGAGGGGCTCGGGGATGGGCACCTGGAGGGGGAGGGGCT
CGGGGATGGGCACCTGGAGGGGGAGGGGCTCGGGGATGGGCACCTGGCGGGGGACGGGCTTGGGGATGGG
CACAGTCCAGGAGGGTTGTGGGCAGTGAGGCCCCACCCTAAGCCTCCATTCCCCCGCCATCCATTCTCAG
AATACGCAGCCCAGCTGGAGCGGCAGCTGCAGTTCTACACGGAGGCTGCCCGGCGCCTGGGCAACGATGG
CAGCAGGGATGCTGCAAAGGAGGCGCTCTATAGGCGGAATCTGGTAGAGAGTGAGCTGCAGCGGCTCCGC
AGGTGAGGAGCCCATGGGGCGGGCAGCCCCCAGAAAGCGGGCAGCAGGCCCCGATACCGGGAAGAGCCGA
CACAGCCACGAACCAGACAAGCAGACAATCAGCGGACAATCGGTTCTGGACTCACCCCTCATCCGGGCCC
CCAGCCCCGCCAGAGCCTCCGTGGCTGCGGGTGTTGGGAACCATGCCTGCCAGCCAGTATGTGCCCCTCA
CCCAGGCCTGGCTGGGCCCTGGAGAGTCCTGTTTGCACAGCCCAGGGGTGTCCGGCCTCTGGCCCGCCCC
GGAGCAGGGAGGGTGGCTGGGGCCAAGCCCCGAGGGCCCCTGCAAGCACTTTACTTCCTGTTCCTCCCCA
GCCTTAACCCCAAAGCCCTCCTGCACCCCAAAGAAGCCACTGAGGCTGGCCGAGCCACACTGTCTCCCCA
GGGGCGTCGACCTGGCCCAGCTGGGTCCCCAGGGCCAGCACATGGAATAAAATAGCCAGGGCCACACTCA

>DEX0356_017.nt.3
GGGAGGGGCTCGGGGATGGGCACCTGGAGGGGGAGGGGCTCGGGGATGGGCACCTGGAGGGGGAGGGGCT
CGGGGATGGGCACCTGGAGGGGGAGGGGCTCGGGGATGGGCACCTGGCGGGGGACGGGCTTGGGGATGGG
CACAGTCCAGGAGGGTTGTGGGCAGTGAGGCCCCACCCTAAGCCTCCATTCCCCCGCCATCCATTCTCAG
AATACGCAGCCCAGCTGGAGCGGCAGCTGCAGTTCTACACGGAGGCTGCCCGGCGCCTGGGCAACGATGG
CAGCAGGGTGAGCTGGTCGCGGGCCGGGTGGGCACTGGGCAGCGGGCAGGGTGGGGCCTGCAGGGACTAC
CTGCTGAATGCCCATCCCCCACAGGATGCTGCAAAGGAGGCGCTCTATAGGCGGAATCTGGTAGAGAGTG
AGCTGCAGCGGCTCCGCAGGTGAGGAGCCCATGGGGCGGGCAGCCCCCAGAAAGCGGGCAGCAGGCCCCG
ATACCGGGAAGAGCCGACACAGCCACGAACCAGACAAGCAGACAATCAGCGGACAATCGGTTCTGGACTC
ACCCCTCATCCGGGCCCCCAGCCCCGCCAGAGCCTCCGTGGCTGCGGGTGTTGGGAACCATGCCTGCCAG
CCAGTATGTGCCCCTCACCCAGGCCTGGCTGGGCCCTGGAGAGTCCTGTTTGCACAGCCCAGGGGTGTCC
GGCCTCTGGCCCGCCCCGGAGCAGGGAGGGTGGCTGGGGCCAAGCCCCGAGGGCCCCTGCAAGCACTTTA
CTTCCTGTTCCTCCCCAGCCTTAACCCCAAAGCCCTCCTGCACCCCAAAGAAGCCACTGAGGCTGGCCGA
GCCACACTGTCTCCCCAGGGGCGTCGACCTGGCCCAGCTGGGTCCCCAGGGCCAGCACATGGAATAAAAT
AGCCAGGGCCACACTCA

>DEX0356_018.nt.1
CAGCCCCTGGAGCTTCTCCGGGCAGAAGAGGAGGTGGGCTGGACTGTCAAAGAAAGTGGCTGGCTTCCGT
GGGCAATCCGAGGTTCCTAGTGGCCCTCTGCAGAAAGCTGAACTCCCAGTTCCCCCAGGCCTCCGTCCCG
CGGCGGACTCCTAGGCTTCCTCCCTCCCTGTCGCCGCCAGCCTCTCGCTCTCCCCGTCTTCCGCCCGCAC
TCTCTGGGCAGGGCTGCGGCCAGGACCCGCCCCCGCGTCCCGCCCGACCCTCCGCCAGGGGTCACTTCCC
CTGTCCAGGTTTCAGCTTCCACATGTGTCAAGCGGCTGGCTCAGCCCAGAGTCCCTGTCTCCCGCCCGCC
GGCCCGAGCCGCCGCCCCTCCCCCGCCTCCCGTGCGCCCGGGACAATCCTCGCCTTGTCTGTGGCGCCGG
CATCTGGAGCTTTCTGTAGCCTCCGGATACGCCTTTTTTTCAGGGCGTAGCCCCAGCCAAGCTGCTCCCC
GCGGCGGCCGCACAGCAGCCCGAGCGCCCCCTTTCCAGAGCTCCCCTCCGGAGCTGGGATCCAGGCGCGT
AGCGGAGATCCCAGGATCCTGGGTGCTGTCTGGGCCCGCTCCCCACCATGACCTCCTCGGGGCCTGGACC
CCGGTTCCTGCTGCTGCTGCCGCTGCTGCTGCCCCCTGCGGCCTCAGCCTCCGACCGGCCCCGGGGCCGA
GACCCGGTCAACCCAGAGAAGCTGCTGGTGATCACTGTGGCCACAGCTGAAACCGAGGGGTACCTGCGTT
TCCTGCGCTCTGCGGAGTTCTTCAACTACACTGTGCGGACCCTGGGCCTGGGAGAGGAGTGGCGAGGGGG
TGATGTGGCTCGAACAGTTGGTGGAGGACAGAAGGTCCGGTGGTTAAAGAAGGAAATGGAGAAATACGCT
GACCGGGAGGATATGATCATCATGTTTGTGGATAGCTACGACGTGATTCTGGCCGGCAGCCCCACAGAGC
TGCTGAAGAAGTTCGTCCAGAGTGGCAGCCGCCTGCTCTTCTCTGCAGAGAGCTTCTGCTGGCCCGAGTG
GGGGCTGGCGGAGCAGTACCCTGAGGTGGGCACGGGGAAGCGCTTCCTCAATTCTGGTGGATTCATCGGT
TTTGCCACCACCATCCACCAAATCGTGCGCCAGTGGAAGTACAAGGATGATGACGACGACCAGCTGTTCT
ACACACGGCTCTACCTGGACCCAGGACTGAGGGAGAAACTCAGCCTTAATCTGGATCATAAGTCTCGGAT
CTTTCAGAACCTCAACGGGGCTTTAGATGAAGTGGTTTTAAAGTTTGATCGGAACCGTGTGCGTATCCGG
AACGTGGCCTACGACACGCTCCCCATTGTGGTCCATGGAAACGGTCCCACTAAGCTGCAGCTCAACTACC
TGGGAAACTACGTCCCCAATGGCTGGACTCCTGAGGGAGGCTGTGGCTTCTGCAACCAGGACCGGAGGAC
ACTCCCGGGGGGGCAGCCTCCCCCCCGGGTGTTTCTGGCCGTGTTTGTGGAACAGCCTACTCCGTTTCTG
CCCCGCTTCCTGCAGCGGCTGCTACTCCTGGACTATCCCCCCGACAGGGTCACCCTTTTCCTGCACAACA
ACGAGGTCTTCCATGAACCCCACATCGCTGACTCCTGGCCGCAGCTCCAGGACCACTTCTCAGCTGTGAA
GCTCGTGGGGCCGGAGGAGGCTCTGAGCCCAGGCGAGGCCAGGGACATGGCCATGGACCTGTGTCGGCAG
GACCCCGAGTGTGAGTTCTACTTCAGCCTGGACGCCGACGCTGTCCTCACCAACCTGCAGACCCTGCGTA
TCCTCATTGAGGAGAACAGGAAGGTGATCGCCCCCATGCTGTCCCGCCACGGCAAGCTGTGGTCCAACTT
CTGGGGCGCCCTGAGCCCCGATGAGTACTACGCCCGCTCCGAGGACTACGTGGAGCTGGTGCAGCGGAAG

CGAGTGGGTGTGTGGAATGTACCATACATCTCCCAGGCCTATGTGATCCGGGGTGATACCCTGCGGATGG
AGCTGCCCCAGAGGGATGTGTTCTCGGGCAGTGACACAGACCCGGACATGGCCTTCTGTAAGAGCTTTCG
AGACAAGGGCATCTTCCTCCATCTGAGCAATCAGCATGAATTTGGCCGGCTCCTGGCCACTTCCAGATAC
GACACGGAGCACCTGCACCCCGACCTCTGGCAGATCTTCGACAACCCCGTCGACTGGAAGGAGCAGTACA
TCCACGAGAACTACAGCCGGGCCCTGGAAGGGGAAGGAATCGTGGAGCAGCCATGCCCGGACGTGTACTG
GTTCCCACTGCTGTCAGAACAAATGTGTGATGAGCTGGTGGCAGAGATGGAGCACTACGGCCAGTGGTCA
GGCGGCCGGCATGAGACTGCACTGCAGGAATGCCTCATGCCGGTCGAGCTCTGCCCTGCTGTGAGCAGCA
GCCACCACACAGCTTCTAAGGCTGCGATGAGCAGCCTGGGTGCAATAGAAGCCACGGGGAAGCCTTGACC
CAGACTTGGGAGGAGTAGGGAGCTTCACAGGACAGGTGACATCTTAGCTGAGACTTGAAGGGCCGGGAGC
AGTGAGTCAGGCCGAGAAGTAGCAAAGAAGACTGTTCTGCACAGAGGAACACAGGTGTGAAGGCTGGAGG
ACGGGAGGAAGGAGGCTGGACTGCCCTGGGGTGGGGAGCGCCATGGCTGGAGGGGCGGGTGGCTCAGTGC
CCGCCGCCCCCAGGATTCAAGGCTGGCTGGAGGCTACGAGAATGTGCCCACCGTGGACATCCACATGAAG
CAGGTGGGGTACGAGGACCAGTGGCTGCAGCTGCTGCGGACGTATGTGGGCCCCATGACCGAGAGCCTGT
TTCCCGGTTACCACACCAAGGCGCGGGCGGTGATGAACTTTGTGGTTCGCTACCGGCCAGACGAGCAGCC
GTCTCTGCGGCCACACCACGACTCATCCACCTTCACCCTCAACGTTGCCCTCAACCACAAGGGCCTGGAC
TATGAGGGAGGTGGCTGCCGCTTCCTGCGCTACGACTGTGTGATCTCCTCCCCGAGGAAGGGCTGGGCAC
TCCTGCACCCCGGCCGCCTCACCCACTACCACGAGGGGCTGCCAACGACCTGGGGCACACGCTACATCAT
GGTGTCCTTTGTCGACCCCTGACACTCAACCACTCCGCCCACAGGGGTTCGTGGGCACAGGGCTTCTGGG
GACTCCCCGCGTGATAAATTATTAATGTTCCGCAGTCTCACTCTGAATAAAGGACAGTTTGTAAGTCTTG
AAATAATAAAATAGAATGGCCTCTGTGGGCCTCACGGAGTGGTGCTTATCTGGGCCGTTCCTTTTATTGT
AGGGTAAATTAAT

>DEX0356_018.nt.2
CAGCCCCTGGAGCTTCTCCGGGCAGAAGAGGAGGTGGGCTGGACTGTCAAAGAAAGTGGCTGGCTTCCGT
GGGCAATCCGAGGTTCCTAGTGGCCCTCTGCAGAAAGCTGAACTCCCAGTTCCCCCAGGCCTCCGTCCCG
CGGCGGACTCCTAGGCTTCCTCCCTCCCTGTCGCCGCCAGCCTCTCGCTCTCCCCGTCTTCCGCCCGCAC
TCTCTGGGCAGGGCTGCGGCCAGGACCCGCCCCCGCGTCCCGCCCGACCCTCCGCCAGGGGTCACTTCCC
CTGTCCAGGTTTCAGCTTCCACATGTGTCAAGCGGCTGGCTCAGCCCAGAGTCCCTGTCTCCCGCCCGCC
GGCCCGAGCCGCCGCCCCTCCCCCGCCTCCCGTGCGCCCGGGACAATCCTCGCCTTGTCTGTGGCGCCGG
CATCTGGAGCTTTCTGTAGCCTCCGGATACGCCTTTTTTTCAGGGCGTAGCCCCAGCCAAGCTGCTCCCC
GCGGCGGCCGCACAGCAGCCCGAGCGCCCCCTTTCCAGAGCTCCCCTCCGGAGCTGGGATCCAGGCGCGT
AGCGGAGATCCCAGGATCCTGGGTGCTGTCTGGGCCCGCTCCCCACCATGACCTCCTCGGGGCCTGGACC
CCGGTTCCTGCTGCTGCTGCCGCTGCTGCTGCCCCCTGCGGCCTCAGCCTCCGACCGGCCCCGGGGCCGA
GACCCGGTCAACCCAGAGAAGCTGCTGGTGATCACTGTGGCCACAGCTGAAACCGAGGGGTACCTGCGTT
TCCTGCGCTCTGCGGAGTTCTTCAACTACACTGTGCGGACCCTGGGCCTGGGAGAGGAGTGGCGAGGGGG
TGATGTGGCTCGAACAGTTGGTGGAGGACAGAAGGTCCGGTGGTTAAAGAAGGAAATGGAGAAATACGCT
GACCGGGAGGATATGATCATCATGTTTGTGGATAGCTACGACGTGATTCTGGCCGGCAGCCCCACAGAGC
TGCTGAAGAAGTTCGTCCAGAGTGGCAGCCGCCTGCTCTTCTCTGCAGAGAGCTTCTGCTGGCCCGAGTG
GGGGCTGGCGGAGCAGTACCCTGAGGTGGGCACGGGGAAGCGCTTCCTCAATTCTGGTGGATTCATCGGT
TTTGCCACCACCATCCACCAAATCGTGCGCCAGTGGAAGTACAAGGATGATGACGACGACCAGCTGTTCT
ACACACGGCTCTACCTGGACCCAGGACTGAGGGAGAAACTCAGCCTTAATCTGGATCATAAGTCTCGGAT
CTTTCAGAACCTCAACGGGGCTTTAGATGAAGTGGTTTTAAAGTTTGATCGGAACCGTGTGCGTATCCGG
AACGTGGCCTACGACACGCTCCCCATTGTGGTCCATGGAAACGGTCCCACTAAGCTGCAGCTCAACTACC
TGGGAAACTACGTCCCCAATGGCTGGACTCCTGAGGGAGGCTGTGGCTTCTGCAACCAGGACCGGAGGAC
ACTCCCGGGGGGGCAGCCTCCCCCCCGGGTGTTTCTGGCCGTGTTTGTGGAACAGCCTACTCCGTTTCTG
CCCCGCTTCCTGCAGCGGCTGCTACTCCTGGACTATCCCCCCGACAGGGTCACCCTTTTCCTGCACAACA
ACGAGGTCTTCCATGAACCCCACATCGCTGACTCCTGGCCGCAGCTCCAGGACCACTTCTCAGCTGTGAA
GCTCGTGGGGCCGGAGGAGGCTCTGAGCCCAGGCGAGGCCAGGGACATGGCCATGGACCTGTGTCGGCAG
GACCCCGAGTGTGAGTTCTACTTCAGCCTGGACGCCGACGCTGTCCTCACCAACCTGCAGACCCTGCGTA
TCCTCATTGAGGAGAACAGGAAGGTGATCGCCCCCATGCTGTCCCGCCACGGCAAGCTGTGGTCCAACTT
CTGGGGCGCCCTGAGCCCCGATGAGTACTACGCCCGCTCCGAGGACTACGTGGAGCTGGTGCAGCGGAAG
CGAGTGGGTGTGTGGAATGTACCATACATCTCCCAGGCCTATGTGATCCGGGGTGATACCCTGCGGATGG
AGCTGCCCCAGAGGGATGTGTTCTCGGGCAGTGACACAGACCCGGACATGGCCTTCTGTAAGAGCTTTCG
AGACAAGGGCATCTTCCTCCATCTGAGCAATCAGCATGAATTTGGCCGGCTCCTGGCCACTTCCAGATAC
GACACGGAGCACCTGCACCCCGACCTCTGGCAGATCTTCGACAACCCCGTCGACTGGAAGGAGCAGTACA
TCCACGAGAACTACAGCCGGGCCCTGGAAGGGGAAGGAATCGTGGAGCAGCCATGCCCGGACGTGTACTG
GTTCCCACTGCTGTCAGAACAAATGTGTGATGAGCTGGTGGCAGAGATGGAGCACTACGGCCAGTGGTCA
GGCGGCCGGCATGAGGTAAGGGCCTGGGCGAGGAGGGCAGAGGCCTTCCCAGAGGAGGAGGTGCAGGATC
TGGTAAACAAACAATGGAATCAGGAAAATGGGAGCAAAGAGGGGTTCCCAGGGGGGAGGCAGACCCTGGG
GTCACAACTGCATTCATTCATGTATTCAAGAAATATTTACTGCTGGGCACAGTGGCTCATGCCTGTAATC

CTAGCACTTTGGGAGGCTGAGGCGGGAGGATTGCTCGAGCAGGGGGAGTTTGAGGCCAGCCTGGG

>DEX0356_018.nt.3

CAGCCCCTGGAGCTTCTCCGGGCAGAAGAGGAGGTGGGCTGGACTGTCAAAGAAAGTGGCTGGCTTCCGT
GGGCAATCCGAGGTTCCTAGTGGCCCTCTGCAGAAAGCTGAACTCCCAGTTCCCCCAGGCCTCCGTCCCG
CGGCGGACTCCTAGGCTTCCTCCCTCCCTGTCGCCGCCAGCCTCTCGCTCTCCCCGTCTTCCGCCCGCAC
TCTCTGGGCAGGGCTGCGGCCAGGACCCGCCCCCGCGTCCCGCCCGACCCTCCGCCAGGGGTCACTTCCC
CTGTCCAGGTTTCAGCTTCCACATGTGTCAAGCGGCTGGCTCAGCCCAGAGTCCCTGTCTCCCGCCCGCC
GGCCCGAGCCGCCGCCCCTCCCCCGCCTCCCGTGCGCCCGGGACAATCCTCGCCTTGTCTGTGGCGCCGG
CATCTGGAGCTTTCTGTAGCCTCCGGATACGCCTTTTTTTCAGGGCGTAGCCCCAGCCAAGCTGCTCCCC
GCGGCGGCCGCACAGCAGCCCGAGCGCCCCCTTTCCAGAGCTCCCCTCCGGAGCTGGGATCCAGGCGCGT
AGCGGAGATCCCAGGATCCTGGGTGCTGTCTGGGCCCGCTCCCCACCATGACCTCCTCGGGGCCTGGACC
CCGGTTCCTGCTGCTGCTGCCGCTGCTGCTGCCCCCTGCGGCCTCAGCCTCCGACCGGCCCCGGGGCCGA
GACCCGGTCAACCCAGAGAAGCTGCTGGTGATCACTGTGGCCACAGCTGAAACCGAGGGGTACCTGCGTT
TCCTGCGCTCTGCGGAGTTCTTCAACTACACTGTGCGGACCCTGGGCCTGGGAGAGGAGTGGCGAGGGGG
TGATGTGGCTCGAACAGTTGGTGGAGGACAGAAGGTCCGGTGGTTAAAGAAGGAAATGGAGAAATACGCT
GACCGGGAGGATATGATCATCATGTTTGTGGATAGCTACGACGTGATTCTGGCCGGCAGCCCCACAGAGC
TGCTGAAGAAGTTCGTCCAGAGTGGCAGCCGCCTGCTCTTCTCTGCAGAGAGCTTCTGCTGGCCCGAGTG
GGGGCTGGCGGAGCAGTACCCTGAGGTGGGCACGGGGAAGCGCTTCCTCAATTCTGGTGGATTCATCGGT
TTTGCCACCACCATCCACCAAATCGTGCGCCAGTGGAAGTACAAGGATGATGACGACGACCAGCTGTTCT
ACACACGGCTCTACCTGGACCCAGGACTGAGGGAGAAACTCAGCCTTAATCTGGATCATAAGTCTCGGAT
CTTTCAGAACCTCAACGGGGCTTTAGATGAAGTGGTTTTAAAGTTTGATCGGAACCGTGTGCGTATCCGG
AACGTGGCCTACGACACGCTCCCCATTGTGGTCCATGGAAACGGTCCCACTAAGCTGCAGCTCAACTACC
TGGGAAACTACGTCCCCAATGGCTGGACTCCTGAGGGAGGCTGTGGCTTCTGCAACCAGGACCGGAGGAC
ACTCCCGGGGGGGCAGCCTCCCCCCCGGGTGTTTCTGGCCGTGTTTGTGGAACAGCCTACTCCGTTTCTG
CCCCGCTTCCTGCAGCGGCTGCTACTCCTGGACTATCCCCCCGACAGGGTCACCCTTTTCCTGCACAACA
ACGAGGTCTTCCATGAACCCCACATCGCTGACTCCTGGCCGCAGCTCCAGGACCACTTCTCAGCTGTGAA
GCTCGTGGGGCCGGAGGAGGCTCTGAGCCCAGGCGAGGCCAGGGACATGGCCATGGACCTGTGTCGGCAG
GACCCCGAGTGTGAGTTCTACTTCAGCCTGGACGCCGACGCTGTCCTCACCAACCTGCAGACCCTGCGTA
TCCTCATTGAGGAGAACAGGAAGGTGATCGCCCCCATGCTGTCCCGCCACGGCAAGCTGTGGTCCAACTT
CTGGGGCGCCCTGAGCCCCGATGAGTACTACGCCCGCTCCGAGGACTACGTGGAGCTGGTGCAGCGGAAG
CGAGTGGGTGTGTGGAATGTACCATACATCTCCCAGGCCTATGTGATCCGGGGTGATACCCTGCGGATGG
AGCTGCCCCAGAGGGATGTGTTCTCGGGCAGTGACACAGACCCGGACATGGCCTTCTGTAAGAGCTTTCG
AGACAAGGGCATCTTCCTCCATCTGAGCAATCAGCATGAATTTGGCCGGCTCCTGGCCACTTCCAGATAC
GACACGGAGCACCTGCACCCCGACCTCTGGCAGATCTTCGACAACCCCGTCGACTGGAAGGAGCAGTACA
TCCACGAGAACTACAGCCGGGCCCTGGAAGGGGAAGGAATCGTGGAGCAGCCATGCCCGGACGTGTACTG
GTTCCCACTGCTGTCAGAACAAATGTGTGATGAGCTGGTGGCAGAGATGGAGCACTACGGCCAGTGGTCA
GGCGGCCGGCATGAGTGGKRAGGTACRAATGGGGCCGGGCCTCACATGAAAGGGGGGCCCGGCCAGGTGA
TCGGGAGTTTGGCCGACAACCTGGCGGCCACAGGCGGCGGGAAGGGTCCGGCCAAAAGTGGCCCAAACCA
CCAGCACAGGGCAAGGGGCCAGACCGGGACGCCAAAAGGGGACACGGGGC

>DEX0356_018.nt.4

GCATTCCAGCCTGGGTGACAGAGCAAGACTCCGTCTCAAAAACCAAAAACAAGAAACAAAAAAGAAGTGC
CACAGCTTGGGGTAGGAACTGACCGCATTTTGGGAGGTGGGGTGGTTGGGACCAGCAGTGCTTGACAAGG
GACATGCAGGGGCCTTATGTTCAGGGACTGGGATACCCTGAGGGAAGAAGGACTTCCCGGGATGAGCTGG
GCAGAATTTGGGGGGCTGTCCTCCCAGCTGACCGCGGTGTCGGTGCCTCCAGGGGGGGTGTGTGGAATGT
ACCATACATCTCCCAGGCCTATGTGATCCGGGGTGATACCCTGCGGATGGAGCTGCCCCAGAGGGATGTG
TTCTCGGGCAGTGACACAGACCCGGACATGGCCTTCTGTAAGAGCTTTCGAGACAAGGTGAGCGCGGGTG
CACGGTCTGGCCTGGGGGCAGTCCCCCAGACTCCAGGCATCGCCTGCATCACGGACACCCCCACCCCACT
ACAGGCATCTTCCTCCATCTGAGCAATCAGCATGAATTTGGCCGGCTCCTGGCCACTTCCAGATACGACA
CGGAGCACCTGCACCCCGACCTCTGGCAGATCTTCGACAACCCCGTCGTGAGTGGGGACCCCACGCCCAG
AGCACACAGGTTCCCCGCCTTGTGCTAAGGAAGACACCAAACGTGATGGCTGCTGCCAGCGCAGGACACT
GAGTGGGAGGGGGCGGTCCCCTGGGGGCTGGGAGCGAGGGGGGCACAGATCTGATGTGCCCCCCACCCTC
TCACAGGACTGGAAGGAGCAGTACATCCACGAGAACTACAGCCGGGCCCTGGAAGGGGAAGGAATCGTGG
AGCAGCCATGCCCGGACGTGTACTGGTTCCCACTGCTGTCAGAACAAATGTGTGATGAGCTGGTGGCAGA
GATGGAGCACTACGGCCAGTGGTCAGGCGGCCGGCATGAGGATTCAAGGCTGGCTGGAGGCTACGAGAAT
GTGCCCACCGTGGACATCCACATGAAGCAGGTGGGGTACGAGGACCAGTGGCTGCAGCTGCTGCGGACGT
ATGTGGGCCCCATGACCGAGAGCCTGTTTCCCGGTTACCACACCAAGGCGCGGGCGGTGATGAACTTTGT
GGTTCGCTACCGGCCAGACGAGCAGCCGTCTCTGCGGCCACACCACGACTCATCCACCTTCACCCTCAAC
GTTGCCCTCAACCACAAGGGCCTGGACTATGAGGGAGGTGGCTGCCGCTTCCTGCGCTACGACTGTGTGA

TCTCCTCCCCGAGGAAGGGCTGGGCACTCCTGCACCCCGGCCGCCTCACCCACTACCACGAGGGGCTGCC
AACGACCTGGGGCACACGCTACATCATGGTGTCCTTTGTCGACCCCTGACACTCAACCACTCTGCCAAAC
CTGCCCTGCCATTGTGCCTTTTTAGGGGGCCTGGCCCCCGTCCTGGGAGTTGGGGGATGGGTCTCTCTGT
CTCCCCACTTCCTGAGTTCATGTTCCGCGTGCCTGAACTGAATATGTCACCTTGCTCCCAAGACACGGCC
CTCTCAGGAAGCTCCCGGAGTCCCCGCCTCTCTCCTCCGCCCACAGGGGTTCGTGGGCACAGGGCTTCTG
GGGACTCCCCGCGTGATAAATTATTAATGTTCCGCAGTCTCACTCTGAATAAAGGACAGTTTGTAAGTCT
TGA

>DEX0356_019.nt.1
GCGGCCGCAGCTCAGAGCCGTGGCCTCTCCGAAGTGGCAGATGGGGCGGGCGCGGCCAGAGCAAGTGCCA
GGCGGGAACAGAGGGACTGGGCGCGCCTCACAACTCACCACCTCGCCCGCTGGTCCTTCCTGGCTCGCCT
GGCTCTGAAGCTGCACCTGGAGGGGAAACCTCAGAACAGTAGGCGGGATTGCCTAGTAAATATCTCCCAT
TCAGGGAGGCCCAGGTCGTGTGACGTCGACAGTTGCTGGGTAGATGAGGCCAACACAGGTTGCAAGAAGA
GGCGGGGTTTAGAGGCGTGAAACTCCGCAGTGCTCAGCCAAGCAGGGAGCAACGCTAGGAAGGGCGGGCA
GAAAGGGCACGCTCTTGTGGGTGACTACAGGTTAGGAGACCGTTGAACCTGGAGGGGCCCTAGGATGGAC
CCCGTGGAAAGATTCAGAGACTGCGCCCTCTCCCTGGCGCCGCCTTCCCCTACACGCGGCGGGTATATTC
TGTTGCAGTTGGCCCAGGACCTGTTTCCAAGACTCTGCCCCCTCGCACTTCCGTCCCTCCTGGTTTTGTA
AAGTGATGCTCATAGGAACCCCCACCCCGCGTGACACTACTCCCAGCTCCTGGCTGACTTCTAGTCTTCT
GGTTGAAGCTGCGCCTTTAGATGACACGACCCTACCCACCCCTGTTTCCAGCGGATGCCCGGGCCTGGAG
TACCTCTTACTGTAACCCATCGCCAAGTGGGCTTTTGAAGGCGCCTGTTCCTTTCTCGCTTTCTTCGGAA
GACCCTTGACCCATCATTCCCCCGACCCCCATAACGGGAGAGCAGAGAAGCCGGTCCCCAGTGTGATGGT
CCTGGTCCAGGCACTAACTGTCCTTTTCTCGGAAAAGGCAGGGGGATGTGGAAAAGAGTCTTGTTCCCTC
CCCTTCGATCTGTGGCTTTCGCTTTCACTTCCTCCTCCGAGAGCGGACAGATCTCTGGGTGCTGGGCGGT
CATGGCGCTACTAGATGTATGCGGAGCCCCCGAGGGCAGCGGCCGGAATCGGCTCTCCCGGTTGCGGGAA
GCGGGCGTCGCTCGGACCCAGGACACTACAGTTTCTCTATGCGATCTCCAGAGCTCGCTTTACCCCGGGG
AATGCAGCCCACAGAATTCTTCCAGTCCCTGGGTGGGGACGGAGAAAGGAACGTTCAGATTGAGATGGCC
CATGGCACCACCACGCTCGCCTTCAAGTTCCAGCATGGAGTGATTGCAGCAGTGGATTCTCGGGCCTCAG
CTGGGTCCTACATTAGTGCCTTACGGGTGAACAAGGTGATTGAGATTAACCCTTACCTGCTTGGCACCAT
GTCTGGCTGTGCAGCAGACTGTCAGTACTGGGAGCGCCTGCTGGCCAAGGAATGCAGGCTGTACTATCTG
CGAAATGGAGAACGTATTTCAGTGTCGGCAGCCTCCAAGCTGCTGTCCAACATGATGTGCCAGTACCGGG
GCATGGGCCTCTCTATGGGCAGTATGATCTGTGGCTGGGATAAGAAGGGTCCTGGACTCTACTACGTGGA
TGAACATGGGACTCGGCTCTCAGGAAATATGTTCTCCACGGGTAGTGGGAACACTTATGCCTACGGGGTC
ATGGACAGTGGCTATCGGCCTAATCTTAGCCCTGAAGAGGCCTATGACCTTGGCCGCAGGGCTATTGCTT
ATGCCACTCACAGAGACAGCTATTCTGGAGGCGTTGTCAATATGTCCACTGTCAGCAGCTGGTCGCGGGC
CACTTGAACGTCCCATGTCTTGCCCCGTGTGCCCTTCCGGGAGCGCCTCCATTGCGGCCGGCGCCGGGGG
GCGTCCCGGGAGGACATGCGGCCCTCCGGGGCACACACTTAGTCCTGGGAAAAAAGGCGTCGTGTGCCAA
CAAAAGGTCGTCTCCACAAGCGGGAAGCATATGGGTCGTTATAAATCACGCGTGTGTGCCGCCAAAATGA
TAATAATAGGAGGAGGCTCGACGCCCTATCCATATAACGGCTCACGCGTCCATACTTCTTATTACTAATT
TATTCTTCGN

>DEX0356_019.nt.2
GCGGCCGCAGCTCAGAGCCGTGGCCTCTCCGAAGTGGCAGATGGGGCGGGCGCGGCCAGAGCAAGTGCCA
GGCGGGAACAGAGGGACTGGGCGCGCCTCACAACTCACCACCTCGCCCGCTGGTCCTTCCTGGCTCGCCT
GGCTCTGAAGCTGCACCTGGAGGGGAAACCTCAGAACAGTAGGCGGGATTGCCTAGTAAATATCTCCCAT
TCAGGGAGGCCCAGGTCGTGTGACGTCGACAGTTGCTGGGTAGATGAGGCCAACACAGGTTGCAAGAAGA
GGCGGGGTTTAGAGGCGTGAAACTCCGCAGTGCTCAGCCAAGCAGGGAGCAACGCTAGGAAGGGCGGGCA
GAAAGGGCACGCTCTTGTGGGTGACTACAGGTTAGGAGACCGTTGAACCTGGAGGGGCCCTAGGATGGAC
CCCGTGGAAAGATTCAGAGACTGCGCCCTCTCCCTGGCGCCGCCTTCCCCTACACGCGGCGGGTATATTC
TGTTGCAGTTGGCCCAGGACCTGTTTCCAAGACTCTGCCCCCTCGCACTTCCGTCCCTCCTGGTTTTGTA
AAGTGATGCTCATAGGAACCCCCACCCCGCGTGACACTACTCCCAGCTCCTGGCTGACTTCTAGTCTTCT
GGTTGAAGCTGCGCCTTTAGATGACACGACCCTACCCACCCCTGTTTCCAGCGGATGCCCGGGCCTGGAG
TACCTCTTACTGTAACCCATCGCCAAGTGGGCTTTTGAAGGCGCCTGTTCCTTTCTCGCTTTCTTCGGAA
GACCCTTGACCCATCATTCCCCCGACCCCCATAACGGGAGAGCAGAGAAGCCGGTCCCCAGTGTGATGGT
CCTGGTCCAGGCACTAACTGTCCTTTTCTCGGAAAAGGCAGGGGGATGTGGAAAAGAGTCTTGTTCCCTC
CCCTTCGATCTGTGGCTTTCGCTTTCACTTCCTCCTCCGAGAGCGGACAGATCTCTGGGTGCTGGGCGGT
CATGGCGCTACTAGATGTATGCGGAGCCCCCGAGGGCAGCGGCCGGAATCGGCTCTCCCGGTTGCGGGAA
GCGGGCGTCGCTCGGACCCAGGACACTACAGTTTCTCTATGCGATCTCCAGAGCTCGCTTTACCCCGGGG
AATGCAGCCCACAGAATTCTTCCAGTCCCTGGGTGGGGACGGAGAAAGGAACGTTCAGATTGAGATGGCC
CATGGCACCACCACGCTCGCCTTCAAGTTCCAGCATGGAGTGATTGCAGCAGTGGATTCTCGGGCCTCAG
CTGGGTCCTACATTAGTGCCTTACGGGTGAACAAGGTGATTGAGATTAACCCTTACCTGCTTGGCACCAT

GTCTGGCTGTGCAGCAGACTGTCAGTACTGGGAGCGCCTGCTGGCCAAGGAATGCAGGCTGTACTATCTG
CGAAATGGAGAACGTATTTCAGTGTCGGCAGCCTCCAAGCTGCTGTCCAACATGATGTGCCAGTACCGGG
GCATGGGCCTCTCTATGGGCAGTATGATCTGTGGCTGGGATAAGAAGGGTCCTGGACTCTACTACGTGGA
TGAACATGGGACTCGGCTCTCAGGAAATATGTTCTCCACGGGTAGTGGGAACACTTATGCCTACGGGGTC
ATGGACAGTGGCTATCGGCCTAATCTTAGCCCTGAAGAGGCCTATGACCTTGGCCGCAGGGCTATTGCTT
ATGCCACTCACAGAGACAGCTATTCTGGAGGCGTTGTCAATATCCCTGAGGAATGGTGGTGACAGTGAAC
GGCCCTTCGGCCGCGCCAGTCGGGACCACCAAGGGGGGGGCGGGCGTCCCCGGGGTGGGAGCGGCACACC
TTTCCGAAAAAGTCGCAAGGTGTCCAGAGGGGGTTGTTTAAACCAGTGGCCGAAGAAAGGGCCCGCTGCG
TACTGGCGTTCTTCAACTTATT

>DEX0356_019.nt.3
GCGGCCGCAGCTCAGAGCCGTGGCCTCTCCGAAGTGGCAGATGGGGCGGGCGCGGCCAGAGCAAGTGCCA
GGCGGGAACAGAGGGACTGGGCGCGCCTCACAACTCACCACCTCGCCCGCTGGTCCTTCCTGGCTCGCCT
GGCTCTGAAGCTGCACCTGGAGGGGAAACCTCAGAACAGTAGGCGGGATTGCCTAGTAAATATCTCCCAT
TCAGGGAGGCCCAGGTCGTGTGACGTCGACAGTTGCTGGGTAGATGAGGCCAACACAGGTTGCAAGAAGA
GGCGGGGTTTAGAGGCGTGAAACTCCGCAGTGCTCAGCCAAGCAGGGAGCAACGCTAGGAAGGGCGGGCA
GAAAGGGCACGCTCTTGTGGGTGACTACAGGTTAGGAGACCGTTGAACCTGGAGGGGCCCTAGGATGGAC
CCCGTGGAAAGATTCAGAGACTGCGCCCTCTCCCTGGCGCCGCCTTCCCCTACACGCGGCGGGTATATTC
TGTTGCAGTTGGCCCAGGACCTGTTTCCAAGACTCTGCCCCCTCGCACTTCCGTCCCTCCTGGTTTTGTA
AAGTGATGCTCATAGGAACCCCCACCCCGCGTGACACTACTCCCAGCTCCTGGCTGACTTCTAGTCTTCT
GGTTGAAGCTGCGCCTTTAGATGACACGACCCTACCCACCCCTGTTTCCAGCGGATGCCCGGGCCTGGAG
TACCTCTTACTGTAACCCATCGCCAAGTGGGCTTTTGAAGGCGCCTGTTCCTTTCTCGCTTTCTTCGGAA
GACCCTTGACCCATCATTCCCCCGACCCCCATAACGGGAGAGCAGAGAAGCCGGTCCCCAGTGTGATGGT
CCTGGTCCAGGCACTAACTGTCCTTTTCTCGGAAAAGGCAGGGGGATGTGGAAAAGAGTCTTGTTCCCTC
CCCTTCGATCTGTGGCTTTCGCTTTCACTTCCTCCTCCGAGAGCGGACAGATCTCTGGGTGCTGGGCGGT
CATGGCGCTACTAGATGTATGCGGAGCCCCCGAGGGCAGCGGCCGGAATCGGCTCTCCCGGTTGCGGGAA
GCGGGCGTCGCTCGGACCCAGGACACTACAGTTTCTCTATGCGATCTCCAGAGCTCGCTTTACCCCGGGG
AATGCAGCCCACAGAATTCTTCCAGTCCCTGGGTGGGGACGGAGAAAGGAACGTTCAGATTGAGATGGCC
CATGGCACCACCACGCTCGCCTTCAAGTTCCAGCATGGAGTGATTGCAGCAGTGGATTCTCGGGCCTCAG
CTGGGTCCTACATTAGTGCCTTACGGGTGAACAAGGTGATTGAGATTAACCCTTACCTGCTTGGCACCAT
GTCTGGCTGTGCAGCAGACTGTCAGTACTGGGAGCGCCTGCTGGCCAAGGAATGCAGGCTGTACTATCTG
CGAAATGGAGAACGTATTTCAGTGTCGGCAGCCTCCAAGCTGCTGTCCAACATGATGTGCCAGTACCGGG
GCATGGGCCTCTCTATGGGCAGTATGATCTGTGGCTGGGATAAGAAGGGTCCTGGACTCTACTACGTGGA
TGAACATGGGACTCGGCTCTCAGGAAATATGTTCTCCACGGGTAGTGGGAACACTTATGCCTACGGGGTC
ATGGACAGTGGCTATCGGCCTAATCTTAGCCCTGAAGAGGCCTATGACCTTGGCCGCAGGGCTATTGCTT
ATGCCACTCACAGAGACAGCTATTCTGGAGGCGTTGTCAATATGTACCACATTGARCGAGATGGTKGGTG
AAAGTAAAGTCGATGTGGACCTGTGCCCGCCGGAAGCCCAAAGTGGG

>DEX0356_020.nt.1
TTAACTAGGTCACTGTAGTACCTCTCCAGCAGGCACCTGGGCACTGCCCTCCATTGTCCCCACTCCTGCA
GCACTTGCCAGGACCCAGTCCCCAGTCCCCAGGGCTTGGTGGTACTTGCAGTGTCTTGTGATG
CTTTGAAGACACCAGCTTTGAAGCAAAAAGTTCTAGGTCTGCTTTCCACTTAAGCCCCATCTTAACTGTA
TGACCTTGGACATGTCACTAGACTTGCTTTAGTGTCTCCATCTGTAAAATGGGGATAAAGATGGACTCTT
CTCTCCCAAATGCCCAGCATGATGTGCAGGGCCTGTCCCCCCAGGAGAACCTGTCTCACCATCCTTCCCC
ACTTCCTCCCACCCCCAGGTGTGCCAGAAATCAGGCGAGATCTCCCTGCTGAAGCAGCAGCTGAAAGAGT
CTCAGGCAGAGCTGGTGCAGAAGGGCAGCGAGCTGGTGGCTCTGCGGGTGGCGCTGCGGGAGGCCCGTGC
TACGCTGCGGGTCAGTGAGGGCCGTGCGCGGGGTCTACAGGAGGCCGCCCGAGCTCGGGAGCTGGAGCTG
GAAGCCTGTTCCCAGGAGCTGCAGCGACACCGCCAGGAAGCTGAGCAGCTGCGGGAGAAAGCTGGGCAGT
TGGATGCTGAGGCGGCCGGACTCCGGGAGCCCCCTGTGCCACCTGCCACCGCTGACCCATTCCTCCTGGC
AGAGAGTGATGAGGCCAAAGTGCAGCGGGCAGCAGCCGGGGTTGGGGGCAGCTTGCGGGCCCAGGTGGAG
CGATTGCGGGTGGAGCTGCAGCGGGAGCGGCGGCGGGGTGAGGAGCAGCGGGACAGCTTTGAGGGGGAGC
GGCTGGCCTGGCAGGCAGAGAAGGAGCAGGTGATCCGCTACCAGAAGCAGCTGCAGCACAACTACATCCA
GATGTACCGGCGCAACCGGCAGCTAGAGCAGGAGCTGCAGCAGCTCAGCCTGGAGCTGGAGGCCCGGGAG
CTCGCTGACCTGGGCCTGGCCGAGCAGGCCCCCTGCATCTGCCTGGAGGAGATCACTGCTACTGAGATCT
AGGGCCCTCAGCAACCAGCTCTGTAGGGAGCTCTGCCAGAGGGGCAGCAGCTGCAGATCCACTTAGGCCC
CAGGGTCCACGGATGGCCCCAAAGGCTGAGGGCCCCAAAGCCACTTGTCTCCTAGGATCCAGGCCTCTGG
GCTTCTGCCAAGAACTCAGGGTGGCCCTATGACTTGGAGGAGCAAGATCAGACCGCTCAAAGGTCCCCGT
GTTCACTGTTACCCAGAGGCTCTTGTTACTACCCACTTCATTCCCCACCGCTGCCAGTGCCACTGCCAAC
CCTGTTCACAGGCGCTTCCAGCCCACTCCAGCCAGGGGAGCAGGGAAGAAGAAGGGGCTCCCTCCTCTTC
ACATTCCCCCCGACCCCAAAGCCAGAGAAAGCCAGATGGCACCAGCTGCTCCGGATGTGCCTGCCCACAT

TGGGGGACAGGGCCGGGCCTGGGCTCGGTTCCCAGGTTTGAGCTCTGCAGCCTCTCTCCTGGAGTGAGGG
GGCTGAAGTCAGACCAAAGGAAGAACTCAGAAATGTCTTGTTTATTTGTGTTTGTGACCAAGCAGCCTCT
CCCTTCACCCAGGTTTATGGCCTCGTTTTCACTTGTATATTTTTCACACTGTAAATTTCTTGTACAAACC
CAAAGAAAAAATTAAAAAAAATTTTTTTGTTTA

>DEX0356_020.nt.2
GCGGCCGCCGCCGCGCCGTGGCCCTTTGTGCTGGAACAATGACCAGCTGCCGCTGGCCCCGGGGCCCCGG
CTGGGGTTGAGGGTGACCACCGCCGCCGCCGCGTCGAGCCGAACCGAGCCCCCAGACACGCAGTCACAGG
GCGTCGGAGGGAGGGCTCCCGACTGCCGCCGAGCAGCCCCTACGTGCGCCCCGGCGGATTCGCGTGACGC
GGGTGCTTGCGCGTGTCCGCGCCCCGGGGCGGAGACGCCCCTCGGACCCTCCCCTGCGCGGTGCCGCCCA
CTCGCACCGGGCGGGGGCTGGCCAGGGCCGGGGGCGCCGCGGAGGCGGAGGCGGGGCTGGGCCGGGCAGC
TCCGAGTGGCGGCGCGCGGGCCAGGGCCGGGGCCGGGGCCGGGGCCGGAGCCGGAGCCGGGGCCGGGCGG
CGCGCCTGTGGAGCGCTGGGGGCGGCCTGGGGCTGAGCATGGAGCAGCGCGGCGGAGGTGAGGGGGGCCG
GGTCTGGGAGGCGGCCTGGGGAGCCCCCAACTTMTGCGCGGTGGAATCCAGCCGGGGGGACGGCGTTCCC
CAGTCGCCCCCACTTTCTGGAAGGGGGGCTGGACTCTGGAGAATGGGTCAGGGGCTCCTGGCTTCCCGCC
GCAGAATGGGAACTCGTGATCGCAGACTTGACCCCGCTTGACTGGTCCTTGGAGTTTGGGTGGGAGGGAG
ACTTTCCCCAACTCTTCAGATCCGTCTCCGGCCCTGAACCAGCTCCAGCCCTCTGCTCTCCTCAGCCCAC
TAAAAGGGGGGTGCTCCTGGGCCAGAGTGATTGCTCAGGGAGTGCTGGGGAGAATGAGGGGTTCCCATCT
CCCACAGCCACAAACGCATTTCTCAAAATGGAGCCTCCACTTGCCCCTCCTTCTCTTTCGGGCTCCAAGA
GCGGGAGCCGGCCTGGCTGGGCCTGGCTTTCCAGACTCCTGGCTCTGGCTCCCTGGATGTGTGGCTCTCA
GGGCCCGGCTCTGTGGCACTTTGGACTTATGGGCCCTCGGCTCCGTGGGGTTGGGGGGCCCTGGCTCACT
GGGTCCCGCTGGCTGTCCCCTTGTCGGCCTGACCCTCGGGCCTCTCCGAGTCTTTGTCTCTTTTGGGTCC
CTGTCTTTCTCTGAGCCTGTCTCCTTCCTTCAATTTTCACTTTCTCCGTGCTCGCTGAAGTGGGGAAATG
CAACAGAGGTGCCAGCCCCCTTGGCCCCCTAAGAAGTCCCTTCCCCACTGTTCACTCCACACCCTGGAGT
GGGCCCAGAGAGGGGAGGGGAGGGGGTTCTCTATCTGCTCTGCTGGGGGCTGGGGTGTGTGGGGTGGTCT
GAGCCCCAGCCCTACAGAGGGGCCTGTGGTCACTCCCTGCTCCTGGCCTACCCCTCCCCCAGCTCCCCGC
AGCTCCTGCCCCAGCGTGGACACGTGGAGCGATTGCGGGTGGAGCTGCAGCGGGAGCGGCGGCGGGGTGA
GGAGCAGCGGGACAGCTTTGAGGGGGAGCGGCTGGCCTGGCAGGCAGAGAAGGAGCAGGTGATCCGCTAC
CAGAAGCAGCTGCAGCACAACTACATCCAGATGTACCGGCGCAACCGGCAGCTAGAGCAGGAGCTGCAGC
AGCTCAGCCTGGAGCTGGAGGCCCGGGAGCTCGCTGACCTGGGCCTGGCCGAGCAGGCCCCCTGCATCTG
CCTGGAGGAGATCACTGCTACTGAGATCTAGGGCCCTCAGCAACCAGCTCTGTAGGGAGCTCTGCCAGAG
GGGCAGCAGCTGCAGATCCACTTAGGCCCCAGGGTCCACGGATGGCCCCAAAGGCTGAGGGCCCCAAAGC
CACTTGTCTCCTAGGATCCAGGCCTCTGGGCTTCTGCCAAGAACTCAGGGTGGCCCTATGACTTGGAGGA
GCAAGATCAGACCGCTCAAAGGTCCCCGTGTTCACTGTTACCCAGAGGCTCTTGTTACTACCCACTTCAT
TCCCCACCGCTGCCAGTGCCACTGCCAACCCTGTTCACAGGCGCTTCCAGCCCACTCCAGCCAGGGGAGC
AGGGAAGAAGAAGGGGCTCCCTCCTCTTCACATTCCCCCCGACCCCAAAGCCAGAGAAAGCCAGATGGCA
CCAGCTGCTCCGGATGTGCCTGCCCACATTGGGGGACAGGGCCGGGCCTGGGCTCGGTTCCCAGGTTTGA
GCTCTGCAGCCTCTCTCCTGGAGTGAGGGGGCTGAAGTCAGACCAAAGGAAGAACTCAGAAATGTCTTGT
TTATTTGTGTTTGTGACCAAGCAGCCTCTCCCTTCACCCAGGTTTATGGCCTCGTTTTCACTTGTATATT
TTTCACACTGTAAATTTCTTGTACAAACCCAAAGAAAAAATTAAAAAAAATTTTTTTGTTTA

>DEX0356_020.nt.3
GCGGCCGCCGCCGCGCCGTGGCCCTTTGTGCTGGAACAATGACCAGCTGCCGCTGGCCCCGGGGCCCCGG
CTGGGGTTGAGGGTGACCACCGCCGCCGCCGCGTCGAGCCGAACCGAGCCCCCAGACACGCAGTCACAGG
GCGTCGGAGGGAGGGCTCCCGACTGCCGCCGAGCAGCCCCTACGTGCGCCCCGGCGGATTCGCGTGACGC
GGGTGCTTGCGCGTGTCCGCGCCCCGGGGCGGAGACGCCCCTCGGACCCTCCCCTGCGCGGTGCCGCCCA
CTCGCACCGGGCGGGGGCTGGCCAGGGCCGGGGGCGCCGCGGAGGCGGAGGCGGGGCTGGGCCGGGCAGC
TCCGAGTGGCGGCGCGCGGGCCAGGGCCGGGGCCGGGGCCGGGGCCGGAGCCGGAGCCGGGGCCGGGCGG
CGCGCCTGTGGAGCGCTGGGGGCGGCCTGGGGCTGAGCATGGAGCAGCGCGGCGGAGGTGAGGGGGGCCG
GGTCTGGGAGGCGGCCTGGGGAGCCCCCAACTTMTGCGCGGTGGAATCCAGCCGGGGGGACGGCGTTCCC
CAGTCGCCCCCACTTTCTGGAAGGGGGGCTGGACTCTGGAGAATGGGTCAGGGGCTCCTGGCTTCCCGCC
GCAGAATGGGAACTCGTGATCGCAGACTTGACCCCGCTTGACTGGTCCTTGGAGTTTGGGTGGGAGGGAG
ACTTTCCCCAACTCTTCAGATCCGTCTCCGGCCCTGAACCAGCTCCAGCCCTCTGCTCTCCTCAGCCCAC
TAAAAGGGGGGTGCTCCTGGGCCAGAGTGATTGCTCAGGGAGTGCTGGGGAGAATGAGGGGTTCCCATCT
CCCACAGCCACAAACGCATTTCTCAAAATGGAGCCTCCACTTGCCCCTCCTTCTCTTTCGGGCTCCAAGA
GCGGGAGCCGGCCTGGCTGGGCCTGGCTTTCCAGACTCCTGGCTCTGGCTCCCTGGATGTGTGGCTCTCA
GGGCCCGGCTCTGTGGCACTTTGGACTTATGGGCCCTCGGCTCCGTGGGGTTGGGGGGCCCTGGCTCACT
GGGTCCCGCTGGCTGTCCCCTTGTCGGCCTGACCCTCGGGCCTCTCCGAGTCTTTGTCTCTTTTGGGTCC
CTGTCTTTCTCTGAGCCTGTCTCCTTCCTTCAATTTTCACTTTCTCCGTGCTCGCTGAAGTGGGGAAATG
CAACAGAGGTGCCAGCCCCCTTGGCCCCCTAAGAAGTCCCTTCCCCACTGTTCACTCCACACCCTGGAGT

GGGCCCAGAGAGGGGAGGGGAGGGGGTTCTCTATCTGCTCTGCTGGGGGCTGGGGTGTGTGGGGTGGTCT
GAGCCCCAGCCCTACAGAGGGGCCTGTGGTCACTCCCTGCTCCTGGCCTACCCCTCCCCCAGCTCCCCGC
AGCTCCTGCCCCAGCGTGGACACCTCTTCCTGGTTCGCAGTGTCTTTTGACTTGGGTGGGAAGGAGGGCG
GTGGATTTAATGGAGGGCATGCGAGAAAGAAGGGGAACAGATCAGGCAAAGGAGTTAGATAATGTCCCAC
CTCTCCCCACACGCCGAAGAGAGAAGGAAACAGGCTCCCCCCACCAATGGGCCCTCAAGCCACCCAGGCT
CCCACCACTCTCATACCTGCGTGCCCACGTTCACAGGCTCCACTGCACCACTCACCTTCCTAGTCACACA
CAAACACAGGAGCTCACACATAAACTCATGAACTTACTGACAAGTCTGTCCAGGCACCGTCTCTCCCACA
CACCACAAATTCACTTGTTAGATCCACATTTATAACTTTCACAGGCTTTCTAATGACCCCTGATCCACAT
ACTCTTCTAAGATATACACATGTGAATTCACACATGAACTCATATGTATGTGAATGAACTGATATATCAA
ACATGCATAGACTCACATGATGTTTTCACACACATTTTCATACACACGTGTTCATTTTCATGCTCCTGTG
CATGTAGTCACACAACAAATAACTAAATGCACACACTCCTACTCCCAGTGCAATTCTCTGTCCACTTCTT
GGCCCCGCTCTGCATGATGACTTCAGCAGATGCCTAAGTCCTGGTCATCAGCGGCCTCGGGACTGGGACT
AGGGAGGGGGTTCCGGGTGGAGGCTTGCAGCTGCTTCCAGGAAGGAGCAGAGCCTGGGGAGGTGGAGCCC
TCTTGCCTGGGCTCTGTGGAAGTGAGGTTTTGGAGCAGCCTCCAGAGGTGAAGGCAGGGCATCTACCTTC
CTCCCCTCAGGGTTAGTTTGGGGACCTTCTGAAGCTGTAGGGTTTCAGAGAGGTTGGCAGTGGGGCAGAG
CCTGTGGAGGCTCCTTCAGGCCTTCAGGCCTCACCGATTGGGTGTGCAGAGGCTAGGAGTTGTCATGGAA
ACTGGTATGAGCCTGCATCCACCTCTGGGAGGCAGGAGACCCCAAAGGACACCTGTGGACTCCCTAAGAG
CCGCGGGTTTCTTATTTAGCCAGATCGACAGAGAGGCAGTTTAGCCGCCTTTCTGTCTGAGAGAAACTGA
GGCCCAGCAGAGGGAGGCAGGGCACTGCCTGGCCTTGGCTCTTTGTTCCCTAATTACCATCTGAGGTGCT
GATTGCACCCCTGAGGTACCTCTGCCCAAACTGCCCTGTAATGAGGCCACCTGGCTGAGAGGACCAGGGA
CCTGTGCTTCAAGCTGGTAGAGAAGGCCCCTCCCTCACCTGAGTCCTGGCCTCTGTGGAGGATGCAGGAT
TGGGCTTGGGCCTGGGTCTTTGCTGCCAGGTGGCCCTGCGGTAATCATCCAGGCCTCCCATGCCTCTCAG
CATCTCTGCTGAGTGAATTCACTCCTCAGCCCTGTTCAGACCAGCATTGGGCAGTGTGGTTTGGGGCCCA
GGGGGAATTGAGATTCATTCCAGCACTTGAAGGGGGAGTCTTGGTGGGGATCAGGGCTTGTTACTGATGT
CAAGGAGGTGTTTGGTCAGGATCAGTGGGTCAGGAGGGGTCTCAAGGTGGAGGGACTCTGACACCACTGC
CCTGCTTACAGGTTGCCTGCGAGGCCGCTGGCCAGGCCTGAGCCTCTGCCACCATGGCCATTGTGCAGAC
TCTGCCAGTGCCACTGGAGCCTGCTCCTGAAGCTGCCACTGCCCCACAAGCTCCAGTCATGGGTAGTGTG
AGCAGCCTTATCTCAGGCCGGCCCTGTCCCGGGGGGCCAGCTCCTCCCCGCCACCACGGCCCTCCTGGGC
CCACCTTCTTCCGCCAGCAGGATGGCCTGCTACGGGGTGGCTATGAGGCACAGGAGCCGCTGTGCCCAGC
TGTGCCCCCTAGGAAGGCTGTCCCTGTCACCAGCTTCACCTACATCAATGAGGACTTCCGGACAGAGTCA
CCCCCCAGCCCAAGCAGTGATGTTGAGGATGCCCGAGAGCAGCGGGCACACAATGCCCACCTCCGCGGCC
CACCACCAAAGCTCATCCCTGTCTCTGGAAAGCTGGAGAAGAACATGGAGAAGATCCTGATCCGCCCAAC
AGCCTTCAAGCCAGTGCTGCCCAAACCTCGAGGGGCTCCGTCCCTGCCTAGCTTCATGGGTCCTCGGGCC
ACCGGGCTGTCTGGGAGCCAGGGCAGCCTGACGCAGCTGTTTGGGGGCCCTGCCTCCTCCTCCTCTCTT
CCTCCTCCTCTTCAGCTGCTGACAAACCCCTGGCATTTAGTGGCTGGGCCAGTGGCTGCCCATCAGGGAC
GCTATCCGACTCTGGCCGAAACTCACTGTCCAGCCTGCCCACCTACAGCACCGGAGGTGCCGAGCCAACC
ACCAGCTCCCCAGGCGGGCACCTGCCTTCCCATGGCTCTGGGCGAGGGGCACTGCCTGGGCCAGCCCGAG
GGGTCCCTACTGGGCCCTCCCACTCAGACAGTGGCCGGTCCTCCTCCAGCAAGAGCACAGGCTCCCTAGG
GGGCCGTGTGGCTGGGGGGCTTTTGGGCAGTGGTACTCGGGCCTCCCCTGACAGCAGCTCCTGTGGGGAG
CGCTCACCACCACCACCCCCGCCTCCACCTCCTTCGGATGAGGCCCTGCTGCACTGTGTCCTGGAAGGAAAGC
TCCGAGACCGGGAGGCAGAGCTTCAGCAGCTGCGGGACAGTCTGGACGAGAATGAGGCTACCATGTGCCA
GGCATACGAGGAGCGGCAGCGGCACTGGCAGCGAGAGCGTGAGGCCCTGCGAGAGGACTGTGCGGCCCAG
GCACAGCGGGCACAGCGGGCCCAACAGCTGCTGCAGCTGCAGGTGTTCCAGCTGCAGCAGGAGAAGCGGC
AATTGCAGGACGACTTCGCACAGCTGCTGCAGGAGCGCGAACAGCTGGAGCGGCGCTGCGCCACCTTGGA
GCGGGAGCGGCGGCGGGGTGAGGAGCAGCGGGACAGCTTTGAGGGGGAGCGGGATCACTGCTACTGAGAT
CTAGGGCCCTCAGCAACCAGCTCTGTAGGGAGCTCTGCCAGAGGGGCAGCAGCTGCAGATCCACTTAGGC
CCCAGGGTCCACGGATGGCCCCAAAGGCTGAGGGCCCCAAAGCCACTTGTCTCCTAGGATCCAGGCCTCT
GGGCTTCTGCCAAGAACTCAGGGTGGCCCTATGACTTGGAGGAGCAAGATCAGACCGCTCAAAGGTCCCC
GTGTTCACTGTTACCCAGAGGCTCTTGTTACTACCCACTTCATTCCCCACCGCTGCCAGTGCCACTGCCA
ACCCTGTTCACAGGCGCTTCCAGCCCACTCCAGCCAGGGGAGCAGGGAAGAAGAAGGGGCTCCCTCCTCT
TCACATTCCCCCCGACCCCAAAGCCAGAGAAAGCCAGATGGCACCAGCTGCTCCGGATGTGCCTGCCCAC
ATTGGGGGACAGGGCCGGGCCTGGGCTCGGTTCCCAGGTTTGAGCTCTGCAGCCTCTCTCCTGGAGTGAG
GGGGCTGAAGTCAGACCAAAGGAAGAACTCAGAAATGTCTTGTTTATTTGTGTTTGTGACCAAGCAGCCT
CTCCCTTCACCCAGGTTTATGGCCTCGTTTTCACTTGTATATTTTTCACACTGTAAATTTCTTGTACAAA
CCCAAAGAAAAAATTAAAAAAAATTTTTTTGTTTA

>DEX0356_021.nt.1
AGCGGATTGATCCGGCGGCGCGGGATTGGCCAGTCCCGGGCCAACACGGGTGGCGGGAGTTTTCAGATGT
GTTGGCTGCGGCCCCGACTTCAGCGGTTGGGCTAGGCTGTGCTGACTGTTTGGGTGGCGGACCCCGGGAG
TAAAACCTGTTGTCGATCCCTCAGCTTCCAGCTTGCGGCTTGCTGAGTGGCCACCTTCTTTCCGGTCCCC

GGAGCTGCGGGGAAAGATGGAACGGCACCAGCCGCGGCTGCACCACCCGGCCCAAGGCTCAGCCGCTGGG
ACTCCCTACCCTTCCTCAGCCTCTCTCCGCGGCTGCCGGGAAAGCAAGATGCCGCGCAGGAAGGGCCCCC
AACACCCTCCGCCGCCCAGTGGCCCCGAGGAGCCTGGGGAGAAGCGCCCCAAGTTTCATTTAAATATTAG
GACACTGACGGATGATATGCTGGACAAATTTGCCAGCATAAGAATTCCAGGGAGCAAGAAAGAGAGACCT
CCACTTCCCAACCTGAAGACTGCATTTGCAAGCAGTGATTGCTCAGCAGCACCTTTAGAGATGATGGAGA
ACTTTCCAAAGCCACTGTCAGAGAATGAACTCTTAGAACTTTTTGAAAAAATGATGGAAGATATGAATTT
AAATGAAGATAAAAAGGCACCATTGCGGGAAAAGGACTTCAGTATCAAAAAAGAAATGGTGATGCAGTAC
ATTAATACTGCTTCTAAGACAGGAAGTCTTAAGAGAAGCCGACAGATCTCACCTCAGGAATTCATTCATG
AGCTGAAAATGGGGTCTGCAGATGAGAGACTTGTCACATGCCTGGAGTCTCTCCGAGTGTCTTTGACCAG
CAATCCTGTGAGTTGGGTGGAAAGCTTTGGACATGAAGGGCTTGGATTATTATTAGACATTTTGGAAAAA
CTGATTAGTGGAAAAATCCAAGAAAAAGTTGTAAAGAAAAATCAACATAAAGTCATACAGTGTCTAAAAG
CCCTGATGAATACGCAGTATGGCTTGGAAAGAATTATGAGTGAGGAGAGGAGCCTTTCCTTATTGGCCAA
AGCCGTGGATCCCAGACACCCCAATATGATGACAGATGTGGTTAAACTTCTCTCTGCGGTATGCATTGTA
GGGGAAGAAAGCATCCTTGAAGAAGTTTTAGAAGCTTTAACTTCAGCTGGTGAAGAAAAAAAAATTGACA
GATTTTTTTGTATTGTGGAAGGCCTCCGGCACAATTCAGTTCAACTGCAAGTAGCTTGTATGCAGCTCAT
CAATGCCCTGGTTACATCTCCTGATGATTTGGATTTCAGGCTTCACATCAGAAATGAATTTATGCGTTGT
GGATTGAAAGAGATATTGCCAAATTTAAAATGCATTAAGAATGATGGCCTGGATATCCAACTTAAAGTCT
TTGATGAGCATAAAGAAGAAGATTTGTTTGAGTTATCCCATCGCCTTGAAGATATTAGAGCTGAACTTGA
TGAAGCATATGATGTTTACAACATGGTGTGGAGCACAGTTAAAGAAACTAGAGCAGAGGGATATTTTATT
TCTATTCTTCAGCATCTTTTGCTGATTCGAAATGATTATTTTATAAGGCAACAATACTTCAAATTAATTG
ATGAGTGTGTATCCCAGATTGTATTGCATAGAGATGGAATGGATCCAGACTTCACATATCGAAAAAGACT
AGATTTAGATTTAACCCAGTTTGTAGACATTTGCATAGATCAAGCAAAACTAGAAGAGTTTGAAGAGAAA
GCATCAGAACTTTACAAGAAATTTGAAAAAGAGTTTACCGACCACCAAGAAACTCAGGCTGAATTGCAGA
AAAAAGAGGCAAAGATTAATGAGCTTCAAGCAGAGCTACAAGCTTTTAAGTCTCAGTTTGGTGCCTTGCC
AGCTGATTGTAATATTCCTTTGCCTCCCTCTAAAGAAGGTGGAACTGGCCACTCAGCACTTCCTCCTCCG
CCTCCACTGCCTTCTGGTGGAGGGGTGCCGCCTCCACCTCCTCCCCCACCACCTCCTCCACTTCCAGGAA
TGCGGATGCCATTCAGTGGTCCTGTGCCTCCACCACCTCCCCTGGGATTCCTTGGAGGACAAAATTCTCC
TCCTCTACCAATCCTGCCATTTGGGTTGAAACCAAAGAAAGAATTTAAACCTGAAATCAGCATGAGAAGA
TTGAATTGGTTAAAGATCAGACCTCATGAAATGACTGAAAACTGTTTCTGGATAAAAGTAAATGAAAATA
AGTATGAAAACGTGGATTTGCTTTGTAAACTTGAGAATACATTTTGTTGCCAACAAAAAGAGAAGAAGA
AGAGGAAGATATTGAAGAGAAGAAATCGATTAAGAAAAAAATTAAAGAACTTAAGTTTTTAGATTCTAAA
ATTGCCCAGAACCTTTCAATCTTCCTGAGCTCTTTTCGGGTGCCATATGAGGAAATCAGAATGATGATAT
TGGAAGTAGATGAAACACGGTTGGCAGAGTCTATGATTCAGAACTTAATAAAGCATCTTCCTGATCAAGA
GCAATTAAATTCATTGTCTCAGTTCAAGAGTGAATATAGCAACTTATGTGAACCTGAGCAGTTTGTGGTT
GTGATGAGCAATGTGAAGAGACTACGGCCACGGCTCAGTGCTATTCTCTTTAAGCTTCAGTTTGAAGAGC
AGGTGAACAACATCAAACCTGACATCATGGCTGTCAGTACTGCCTGCGAAGAGATAAAGAAGAGCAAAAG
CTTTAGCAAGTTGCTGGAACTTGTATTGCTAATGGGAAACTACATGAATGCTGGCTCCCGGAATGCTCAA
ACCTTCGGATTTAACCTTAGCTCTCTCTGTAAACTAAAGGACACAAAATCAGCAGATCAGAAAACAACGC
TACTTCATTTCCTGGTAGAAATATGTGAAGAGAAGTACCCTGATATACTGAATTTTGTGGATGATTTGGA
ACCTTTAGACAAAGCTAGTAAAGTCTCTGTAGAAACGCTGGAAAAGAATTTGAGGCAGATGGGAAGGCAG
CTTCAACAGCTTGAGAAGGAATTGGAAACCTTTCCCCCTCCTGAGGACTTGCATGACAAGTTTGTGACAA
AGATGTCCAGATTTGTTATCAGTGCAAAAGAACAATATGAGACACTTTCGAAGTTACACGAAAACATGGA
AAAGTTATACCAGAGTATAATAGGATACTATGCCATTGATGTGAAGAAGGTGTCTGTGGAAGACTTTCTT
ACTGACCTGAATAACTTCAGAACCACATTCATGCAAGCAATAAAGGAGAATATCAAAAAAAGAGAAGCAG
AGGAAAAAGAAAAACGTGTCAGAATAGCTAAAGAATTAGCAGAGCGAGAAAGACTCGAACGCCAACAAAA
GAAAAAGCGTTTATTAGAAATGAAGACTGAGGGTGATGAGACAGGAGTGATGGATAATCTGCTGGAGGCC
TTGCAGTCCGGGGCTGCCTTCCGCGACAGAAGAAAAAGGACACCGATGCCAAAAGATGTTCGGCAGAGTC
TCAGTCCAATGTCTCAGAGGCCTGTTCTGAAAGTTTGTAACCATGAAAATCAGAAAGTGCAGTTGACAGA
AGGGTCACGTTCACACTACAATATCAATTGCAACTCAACAAGGACTCCAGTCGCCAAGGAGCTTAATTAT
AATCTAGACACTCATACGTCTACTGGGAGGATCAAGGCAGCTGAGAAGAAGGAAGCGTGTAATGTAGAAA
GCAACAGAAAAAAGGAAACGGAACTTCTTGGCTCTTTTTCTAAAAATGAATCAGTTCCCGAAGTTGAAGC
CCTGCTGGCAAGATTACGAGCTTTATAAGTTAAACTGGTTTTTAAAAAAAATGATTAAGCCAAATATAAA
GCCATGCTCTAAACTATAACACTTGAAAAAATTGCTTTTATGTAAGTGACTTTATATAGTTTTAAATTAT
GATATATATTAGAAAAATAAAGCTAAATATATGAATGCAATGCTTTTTCTATGTACTGGAGGTTTTGGTT
TATTCAAGATTGTACCGGGCTTTAATCCTTTTTTTTTTTTTTTTTGTCTCCTGGTATTCCTCAGTTCTTTA
TTTGGTGGTTAAATTATACATATTGCTTTAGAGAGCAGGTAGGTGGCCATGTGTTCAGCAGTGTGTCCTT
AAGAAAATACCATCTTTCTAAGCCACTGGAATTTTTACTTTACTATTTTTAACATTAATGGATGTCAGGT
CATCAACCTCAAGTCTTTACATATCCATGGTATATTCCATATATATTGTTTATATAGGCCCAAGTTTCTC
CTTAATTGGGATCTATATACTACCAGCACAACATCAAAAACATGTAATTGAATACATCAGAGCTATATAT
GTAAGGAAATGACTGGTGACCCCATTATCATCATTGTTGAATTCATGTTAAGTAGACCCTCTAGGGGACC

ATAAGGCAATTGAGCACATAACGAAAAATGATGCAATAAGAATGTATGCACTCTCTTTGCCAAATGCATG
TGCTTTTGTGTAACGTGGATGTAAACAGAATTGCAGTGCTGCCGAAATTCTTGATCTTGGCTAAGAGAGT
ATTTTTCCCCTTGTAATTATGACTCTGAGATAAAATTGCCATTTTGAAATTTCCAAAGTAACAACTTTTT
TTATTTTATGAATAAACTTGGGATTGCAATTTCTCTGATCTGACAATCAATAACTTTAACAAAGATCTAA
ATAAGTGTTTCAAGGAAAGTTTTCCTAAGCAAATGTAATATTACCTCATTTGGGCATCATTACTCTGTTA
ATTCTATATCAAAGGAAATAAACTTGCTACTTGCACTAAATGAAAAAACACGAACAAAAGGAAGAACCAG
AAAAAAACAGAAAAAACAAGAAAATACGAAAAAAGGAAACAGAAAGCAACAAAGAAAGAACAAAGCCGGA
GAAGAAGAAAGACAAAAAAAAAAGAGGCCCAGATCATAACCCGAGAGAAAACGAGACTCGAGATTAGAGA
AAGAAGTAAATAAACCGCGGGGTAATGAGAGAAAGAAGCAGAAGCACGAGGAGAGAAACACCAAGAGAGA
ATAGAAAACATCAACAATAAAN

>DEX0356_022.nt.1
ACGTCCACCCCGAATCCCTGCTTAAAGGCCTTGCTTTCTTGTCTAACGCCGCAACCAGTCCTCTGAGTTG
CCAACGTCTTTCTTCTTGTCTCGACGCCCCGTCGTCCGGCCACAGCGATTCTCTGCTTAGCAGGATCGGT
CCACAGCGGGACGTGAGTCCCTTTCCTCCTCGCGGCTTACCGCCTCTCTCCGCCTAGTGCCAGGTGCTAA
TAAAGTTGTTGTTTCAAATGCGGCCAGGAACATCGCGAGCGGGGACCAATCAGAGAGTAGCTTTGCCTCT
ATAACGGCGCGAGAGTGAGACGTCATCGGTGAGCGACTAACGCTAGAAACAGTGGTGCGCGGAGAGGAGA
GGCCTCGGGATGTCTCTGGCAGATGAGCTCTTAGCTGATCTCGAAGAGGCAGCAGAAGAGGAGGAAGGAG
GAAGCTATGGGGAGGAAGAAGAGGAGCCAGCGATCGAGGATGTGCAGGAGGAGACACAGCTGGATCTTTC
CGGGGATTCAGTCAAGACCATCGCCAAGCTATGGGATAGTAAGATGTTTGCTGAGATTATGATGAAGATT
GAGGAGTATATCAGCAAGCAAGCCAAAGCTTCAGAAGTGATGGGACCAGTGGAGGCCGCGCCTGAATACC
GCGTCATCGTGGATGCCAACAACCTGACCGTGGAGATCGAAAACGAGCTGAACATCATCCATAAGTTCAT
CCGGGATAAGTACTCAAAGAGATTCCCTGAACTGGAGTCCTTGGTCCCCAATGCACTGGATTACATCCGC
ACGGTCAAGGAGCTGGGCAACAGCCTGGACAAGTGCAAGAACAATGAGAACCTGCAGCAGATCCTCACCA
ATGCCACCATCATGGTCGTCAGCGTCACCGCCTCCACCACCCAGGGGCAGCAGCTGTCGGAGGAGGAGCT
GGAGCGGCTGGAGGAGGCCTGCGACATGGCGCTGGAGCTGAACGCCTCCAAGCACCGCATCTACGAGTAT
GTGGAGTCCCGGATGTCCTTCATCGCACCCAACCTGTCCATCATTATCGGGGCATCCACGGCCGCCAAGA
TCATGGGTGTGGCCGGCGGCCTGACCAACCTCTCCAAGATGCCCGCCTGCAACATCATGCTGCTCGGGGC
CCAGCGCAAGACGCTGTCGGGCTTCTCGTCTACCTCAGTGCTGCCCCACACCGGCTACATCTACCACAGT
GACATCGTGCAGTCCCTGCCACCGGATCTGCGGCGGAAAGCGGCCCGGCTGGTGGCCGCCAAGTGCACAC
TGGCAGCCCGTGTGGACAGTTTCCACGAGAGCACAGAAGGGAAGGTGGGCTACGAACTGAAGGATGAGAT
CGAGCGCAAATTCGACAAGTGGCAGGAGCCGCCGCCTGTGAAGCAGGTGAAGCCGCTGCCTGCGCCCCTG
GATGGACAGCGGAAGAAGCGAGGCGGCCGCAGGTGAGGGGCCCTGGGGGTCCGGTAGGCATGGGGGTCAT
GGAGGGGAGAAGCCGGCGTCCTCCTCCCAGCCGACTCCCTGGCGCCGCCCACCCACCCGTCCCCGTACCG
CAAGATGAAGGAGCGGCTGGGGCTGACGGAGATCCGGAAGCAGGCCAACCGTATGAGCTTCGGAGAGGTC
AGACTCCCAGAGCGCCCTCCTCAACCCCACAGCCAGCCAGCCGCCACCGCCCTCTGCCTCCTGCCACCGC
CCCTCCTCTCGTCCTGTGGCCCTGGCTCATGTCTAGGGCGCTGCCCCAGCCTCCCCCCCCGCCGGCCTCT
ATTCTCGTTTCCATCCATTCAGCCCCAAAGCGACCCTCGCGGCCCTTGGAGCCTGTGTCTCCGCTGCTTA
GAGCCCCCGCGGCTTCCCATCGCCCCGGGCTCCTTGGCCGGTTCCTCCCTGCCCAGAGGCTCCTTAGTGC
CCTGCTGCACGGCCGCCCCGTCCCTGGGCCCGGCCAGTCTCCTCTGTTATCCCAGCGTCATCCCCTTGGT
CCTGCAGGACCGAAACTCAGAGGCCACCTCATCCTATTAAACCTGTTCTGGTTCCTGACATCCCCCGACC
CACACGAATCGAGGAGGACGCCTACCAGGAGGACCTGGGATTCAGCCTGGGCCACCTGGGCAAGTCGGGC
AGTGGGCGTGTGCGGCAGACACAGGTAAACGAGGCCACCAAGGCCAGGATCTCCAAGACGCTGCAGCGGA
CCCTGCAGAAGCAGAGCGTCGTATATGGCGGGAAGTCCACCATCCGCGACCGCTCCTCGGGCACGGCCTC
CAGCGTGGCCTTCACCCCACTCCAGGGCCTGGAGATTGTGAACCCACAGGCGGCAGAGAAGAAGGTGGCT
GAGGCCAACCAGAAGTATTTCTCCAGCATGGCTGAGTTCCTCAAGGTCAAGGGCGAGAAGAGTGGCCTTA
TGTCCACCTGAATGACTGCGTGTGTCCAAGGTGGCTTCCCACTGAAGGGACACAGAGGTCCAGTCCTTCT
GAAGGGCTAGGATCGGGTTCTGGCAGGGAGAACCTGCCCTGCCACTGGCCCCATTGCTGGGACTGCCCAG
GGAGGAGGCCTTGGAAGAGTCCGGCCTGGCCTCCCCCAGGACCGAGATCACCGCCCAGTATGGGCTAGAG
CAGGTCTTCATCATGCCTTGTCTTTTTTAACTGAGAAAGGAGATTTTTTGAAAAGAGTACAATTAAAAGG
ACATTGTCAAGA

>DEX0356_023.nt.1
CGCCGCCCCTATGCTGGGACCAGGCTCCCAGTGGGCAGCCCTGGSCTCTTCCTTAACGCTGCAGCTGCGA
CTCCACGTACTCCCTCAGTGACTGGAGCGACGGAGACCGTCACGCCCTCAGAGGCCCCGGTGCTGGCAGC
TGAGCCCGAGGCTGACAAGGGCACTGTGTTGGCACTCACTGAAAATAACTTCGATGACACCATTGCAGAA
GGAATAACCTTCATCAAGTTTTATGCTCCATGGTGTGGTCATTGTAAGACTCTGGCTCCTACTTGGGAGG
AACTCTCTAAAAAGGAATTCCCTGGTCTGGCGGGGGTCAAGATCGCCGAAGTAGACTGCACTGCTGAACG
GAATATCTGCAGCAAGTATTCGGTACGAGGCTACCCCACGTTATTGCTTTTCCGAGGAGGGAAGAAAGTC
AGTGAGCACAGTGGAGGCAGAGACCTTGACTCGTTACACCGCTTTGTCCTGAGCCAAGCGAAAGACGAAC

TTTAGGAACACAGTTGGAGGTCACCTCTCCTGCCCAGCTCCCGCACCCTGCGTTTAGGAGTTCAGTCCCA
CAGAGGCCACTGGGTTCCCAGTGGTGGCTGTTCAGAAAGCAGAACATACTAAGCGTGAGGTATCTTCTTT
GTGTGTGTGTTTTCCAAGCCAACACACTCTACAGATTCTTTATTAAGTTAAGTTTCTCTAAGTAAATGTG
TAACTCATGGTCACTGTGTAAACATTTTCAGTGGCGATATATCCCCTTTGACCTTCTCTTGATGAAATTT
ACATGGTTTCCTTTGAGACTAAAATAGCGTTGAGGGAAATGAAATTGCTGGACTATTTGTGGCTCCTGAG
TTGAGTGATTTTGGTGAAAGAAAGCACATCCAAAGCATAGTTTACCTGCCCACGAGTTCTGGAAAGGTGG
CCTTGTGGCAGTATTGACGTTCCTCTGATCTTAAGGTCACAGTTGACTCAATACTGTGTTGGTCCGTAGC
ATGGAGCAGATTGAAATGCAAAAACCCACACCTCTGGAAGATACCTTCACGGCCGCTGCTGGAGCTTCTG
TTGCTGTGAATACTTCTCTCAGTGTGAGAGGTTAGCCGTGATGAAAGCAGCGTTACTTCTGACCGTGCCT
GAGTAAGAGAATGCTGATGCCATAACTTTATGTGTCGATACTTGTCAAATCAGTTACTGTTCAGGGGATC
CTTCTGTTTCTCACGGGGTGAAACATGTCTTTAGTTCCTCATGTTAACACGAAGCCAGAGCCCACATGAA
CTGTTGGATGTCTTCCTTAGAAAGGGTAGGCATGGAAAATTCCACGAGGCTCATTCTCAGTATCTCATTA
ACTCATTGAAAGATTCCAGTTGTATTTGTCACCTGGGGTGACAAGACCAGACAGGCTTTCCCAGGCCTGG
GTATCCAGGGAGGCTCTGCAGCCCTGCTGAAGGGCCCTAACTAGAGTTCTAGAGTTTCTGATTCTGTTTC
TCAGTAGTCCTTTTAGAGGCTTGCTATACTTGGTCTGCTTCAAGGAGGTCGACCTTCTAATGTATGAAGA
ATGGGATGCATTTGATCTCAAGACCAAAGACAGATGTCAGTGGGCTGCTCTGGCCCTGGTGTGCACGGCT
GTGGCAGCTGTTGATGCCAGTGTCCTCTAACTCATGCTGTCCTTGTGATTAAACACCTCTATCTCCCTTG
GGAATAAGCACATACAGGCTTAAGCTCTAAGATAGATAGGTGTTTGTCCTTTTACCATCGAGCTACTTCC
CATAATAACCACTTTGCATCCAACACTCTTCACCCACCTCCCATACGCAAGGGGATGTGGATACTTGGCC
CAAAGTAACTGGTGGTAGGAATCTTAGAAACAAGACCACTTATACTGTCTGTCTGAGGCAGAAGATAACA
GCAGCATCTCGACCAGCCTCTGCCTTAAAGGAAATCTTTATTAATCACGTATGGTTCACAGATAATTCTT
TTTTTAAAAAAACCCAACCTCCTAGAGAAGCACAACTGTCAAGAGTCTTGTACACACAACTTCAGCTTTG
CATCACGAGTCTTGTATTCCAAGAAAATCAAAGTGGTACAATTTGTTTGTTTACACTATGATACTTTCTA
AATAAACTCTTTTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAACGACGAAAAAAGAAAAAAGGGCGGTC
CGAATCCCCGGGCCAGTACGTCCCGTGTGTAAAGGGGCCAAAAGGACGAATTATAGACAGGCAGGGGCGC
GTTAACACGGTTGTGCGGCAAAACGTTATGGGAATTTGGAAGGGACCCTCTTGGTGGGTGACATATTGAC
ACATACCTAGGAATTAGCTTAGGATAATAAAATTTAGGGTTTAATGGTTAACAAGGGATTTTCTGGGTTA
AAATTTGCTGTCGAAATTTTTTTGAAAATTTTTTTCTGGACTTTTAAATAGGGGTT

>DEX0356_023.nt.2
ACGGACCAACACAGTATTGAGTCAACTGTGACCTTAAGATCAGRGGAACGTCAATACTGCCACAAGGCCA
CCTTTCCAGAACTCGTGGGCAGGTAWACTATGCTTTGGAGCAGATTGAAATGCAAAAACCCACACCTCTG
GAAGATACCTTCACGGCCGCTGCTGGAGCTTCTGTTGCTGTGAATACTTCTCTCAGTGTGAGAGGTTAGC
CGTGATGAAAGCAGCGTTACTTCTGACCGTGCCTGAGTAAGAGAATGCTGATGCCATAACTTTATGTGTC
GATACTTGTCAAATCAGTTACTGTTCAGGGGATCCTTCTGTTTCTCACGGGGTGAAACATGTCTTTAGTT
CCTCATGTTAACACGAAGCCAGAGCCCACATGAACTGTTGGATGTCTTCCTTAGAAAGGGTAGGCATGGA
AAATTCCACGAGGCTCATTCTCAGTATCTCATTAACTCATTGAAAGATTCCAGTTGTATTTGTCACCTGG
GGTGACAAGACCAGACAGGCTTTCCCAGGCCTGGGTATCCAGGGAGGCTCTGCAGCCCTGCTGAAGGGCC
CTAACTAGAGTTCTAGAGTTTCTGATTCTGTTTCTCAGTAGTCCTTTTAGAGGCTTGCTATACTTGGTCT
GCTTCAAGGAGGTCGACCTTCTAATGTATGAAGAATGGGATGCATTTGATCTCAAGACCAAAGACAGATG
TCAGTGGGCTGCTCTGGCCCTGGTGTGCACGGCTGTGGCAGCTGTTGATGCCAGTGTCCTCTAACTCATG
CTGTCCTTGTGATTAAACACCTCTATCTCCCTTGGGAATAAGCACATACAGGCTTAAGCTCTAAGATAGA
TAGGTGTTTGTCCTTTTACCATCGAGCTACTTCCCATAATAACCACTTTGCATCCAACACTCTTCACCCA
CCTCCCATACGCAAGGGGATGTGGATACTTGGCCCAAAGTAACTGGTGGTAGGAATCTTAGAAACAAGAC
CACTTATACTGTCTGTCTGAGGCAGAAGATAACAGCAGCATCTCGACCAGCCTCTGCCTTAAAGGAAATC
TTTATTAATCACGTATGGTTCACAGATAATTCTTTTTTTAAAAAAACCCAACCTCCTAGAGAAGCACAAC
TGTCAAGAGTCTTGTACACACAACTTCAGCTTTGCATCACGAGTCTTGTATTCCAAGAAAATCAAAGTGG
TACAATTTGTTTGTTTACACTATGATACTTTCTAAATAAACTCTTTTTTTTTTAAAAAAAAAAAAAAAAAA
AAAAAAAACGACGAAAAAAGAAAAAAGGGCGGTCCGAATCCCCGGGCCAGTACGTCCCGTGTGTAAAGGG
GCCAAAAGGACGAATTATAGACAGGCAGGGGCGCGTTAACACGGTTGTGCGGCAAAACGTTATGGGAATT
TGGAAGGGACCCTCTTGGTGGGTGACATATTGACACATACCTAGGAATTAGCTTAGGATAATAAAATTTA
GGGTTTAATGGTTAACAAGGGATTTTCTGGGTTAAAATTTGCTGTCGAAATTTTTTTGAAAATTTTTTTC
TGGACTTTTAAATAGGGGTT

>DEX0356_023.nt.3
TGATGGTGACTCTAGAGTATCCCSTACGAGAGAAGAGACGTCAAGACCAAAGACAGATGTCAGTGGGCTG
CTCTGGCCCTGGTGTGCACGGCTGTGGCAGCTGTTGATGCCAGTGTCCTCTAACTCATGCTGTCCTTGTG
ATTAAACACCTCTATCTCCCTTGGGAATAAGCACATACAGGCTTAAGCTCTAAGATAGATAGGTGTTTGT
CCTTTTACCATCGAGCTACTTCCCATAATAACCACTTTGCATCCAACACTCTTCACCCACCTCCCATACG
CAAGGGGATGTGGATACTTGGCCCAAAGTAACTGGTGGTAGGAATCTTAGAAACAAGACCACTTATACTG

TCTGTCTGAGGCAGAAGATAACAGCAGCATCTCGACCAGCCTCTGCCTTAAAGGAAATCTTTATTAATCA
CGTATGGTTCACAGATAATTCTTTTTTTAAAAAAACCCAACCTCCTAGAGAAGCACAACTGTCAAGAGTC
TTGTACACACAACTTCAGCTTTGCATCACGAGTCTTGTATTCCAAGAAAATCAAAGTGGTACAATTTGTT
TGTTTACACTATGATACTTTCTAAATAAACTCTTTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAACGA
CGAAAAAAGAAAAAAGGGCGGTCCGAATCCCCGGGCCAGTACGTCCCGTGTGTAAAGGGGCCAAAAGGAC
GAATTATAGACAGGCAGGGGCGCGTTAACACGGTTGTGCGGCAAAACGTTATGGGAATTTGGAAGGGACC
CTCTTGGTGGGTGACATATTGACACATACCTAGGAATTAGCTTAGGATAATAAAATTTAGGGTTTAATGG
TTAACAAGGGATTTTCTGGGTTAAAATTTGCTGTCGAAATTTTTTTGAAAATTTTTTTCTGGACTTTTAA
ATAGGGGTT

>DEX0356_024.nt.1
CCAGGCGACTCTAGAGGATCCGGGCTCCACTTGGTTGGGTTGCAGCCGGAGCCGCCCAGCTCACCGAGAG
CCTAGTTCCGGCCAGGGTCGCCCCGGCAACCACGAGCCCAGCCAATCAGCGCCCCGGACTGCACCAGAGC
CATGGTCGGCAGAAGAGCACTGATCGTACTGGCTCACTCAGAGAGGACGTCCTTCAACTATGCCATGAAG
GAGGCTGCTGCAGCGGCTTTGAAGAAGAAAGGATGGGAGGTGGTGGAGTCGGACCTCTATGCCATGAACT
TCAATCCCATCATTTCCAGAAAGGACATCACAGGTAAACTGAAGGACCCTGCGAACTTTCAGTATCCTGC
CGAGTCTGTTCTGGCTTATAAAGAAGGCCATCTGAGCCCAGATATTGTGGCTGAACAAAAGAAGCTGGAA
GCCGCAGACCTTGTGATATTCCAGGTATGGGGGGACATCGGAAGGGGTGTCAGGGACATTTGCGTGCTTA
TTGTCCTAGACATGTACCTAATTAGCTATGAGATCTTAAACAGTTACCCACTTTACTGCATTCTCTGCAA
ATAAGGGTGATTACTTTAAAGGTGCAGTATTACATGGATTTCACAGATATTGGGAATC

>DEX0356_025.nt.1
TTCGTCTCGCTTTGTTCTGGTTCTTCCTCGTGTCGGTCGTTATTGTTCTTCTTCCGCTTCGTTCTCATTC
GTTGGTCCTTTCCCTTGCTCTCTTCCTCCTTTTCCTCCTCCCTCTTCGTGGTTCTCTTCCGGCGCTGTGG
TCTTGTTCGTTTCTCTCGGTCTGTCCTGGCTTCCGTCTTGTTGGCTCTTCTCCTTCTGTCTTCTTGCGTC
CGGTTTCCTGTTGCTTGTCTGTCTTTCTCGCTTCTGCTCGTCATCTGTTTCTCCTTGTTCTTGCTTTTTC
TCTCTCCGGTCTCGCCTTCTTTCCTTGTCTCCTCCTCCCCTTTTCTTCTCTTTGCGTGTGCTTGTCTCGC
TCGCTCTGTGTTCTTCTGTCTCTGCTTTTGTCGTGTCCGCTTGTCTCTTGTCTTCTTCGGCTTGCTTTTT
CTCTTCTCTCCTCTCCGTTCGCTGCTGTTTTCTGTTCTCCGTGCCTCTGTCCCTTTTGTCTTCTTCGTCT
TCTTCGCTTCCTTCCGTTCCCTCCGCTCCTCTTCCTCCGTTCCTCTCCTCTCGTCGTTCCTCCCGCTCTC
CCCTTTCGCTGCTGCTCTGGCTTCCCTCGCTGGCCGTCCTCCCTCTTCGCCTCCCTCTCCTTCCCTCTGT
CGTTTCTCGTTGTTGTTCTTGTGTGTTGTTGTGTGTGCTCGTGTTGTTCTGGTTCTTGGTTGGTGGTTGT
GTGGTGTGTGCGCTGTGTGTGTTGTTTGTTGTGTTTGTGCGCTCGTGGTGTACTGCTGAGAAGTCACACC
ACCAGCGTACGTCGTTTAACCGGCTATCGTGGGCGCATCCCCCGAGGGTCTCCGAGCCGGGCGGTCCGGC
GGCTGTTCCCGCCTTCTTTTTTTTTGCTCCTTGGGCCCTGTCCAAACGGTCCCGTACTTGGCCCTCGAGGG
AACTTTGGCGCCCCCTTTTTTTTTTTTGCATGTCCACCTTCGCTTTTATTGAGTAGTTAGTGTTCTCAAGC
TGGCTCACACCCATGCGGTCAGCTGGGGCCTAGGGTGGCTCTTTGCAAAGCTGAGGGGCAAGCTAAGGAA
GCCAGGCAGGTCAGGGGCCCTTTCGGCCTTCTCAAGCCTCCACCTGAGTTCTCGTCAATGCCAGTCTCCC
TGGTATGATTGGGGACATTATCAGAGAAACATCTAATAGCGCACATCTGGGCACCCACACTCTGCTTCAG
TTGCATCCATCCTCCCACCCCAAATTCAACTCCTGACCCAATACAAAAGACTTTTTTAACCAGGATTTCT
TCTTGCAGGAAAGCTGACTTGGAAACACGGGGAGGTGGCAGGGAGGGGCAAAAAGGACTCTGGCAAGCAG
ATCCACTTGTCTGGGTCCCTGCAGTGAAGAACCCAAGATCCAGGTACCTCAGCCTGGAAGAAACCGTGCA
CTGCAGGTCTTCCCTTCTATCAAAACTAAAAGCAAGGAACAGAAGTCAATTGAAACCCCAGGTCATCCTA
CCCACCTTCTGGCAGTCTCTGGATAATGATCTAAGGAAATAAGCTTACATTTATAATTAGAGGATAATTT
GGAGAATTTATACTATTTACACGGACCACCCTGCAAAGATCAGGGAAGGTAAGAGACTGCTGGGCAGAGA
ATGGAAACAGGCTGGGGGGTTCCTAATCTCTTTTCACCATTGCCTTTGTTGTTCCCACCCTTCAGCTCCT
GGCAGGGACAGCAGAGGAATCAGATACTCTTGGGGAACACAAGGTCCCAGCAGTGGCTAGGGTGTGACTG
CTACTTTTGCATAATCAGGCAGCAGGGAGCTATGAGGAGCAGAACAGTCCCTGCCTGCTGTCCTCACTCA
GAGGCTTGGGGTGCACTGAGCCTTGGAGTGCCCGGGATGGTGGACAGCTTGAGCACAGGGGCAAGGGCAG
GGCTCTACTGTAGCTCAGGAATAAACTGGGACCAGTTGATGTTGTCAGGGCCCAGTGGGTCCTCGTCCAG
GCCAGGAAAGCTGATGTCCAGCAGGATCTTGCTGAGGCTGTCATTCATTGTGTCCAGGACCAGGCCTTCT
GTCAGAGAACGATTGGCTGCAAGGCCAGAAACCTGTGGCTCCGGGGAGCCTGGCTTGGGGACGTCTATAT
CTGAGGGAGAAGAGTTGCCAAAGGGGACGGAGATGAGGTCTAAGGGTTCTGAACTGAGGAGCCTTTGCGG
TGATTCAAGGGGGGGAGCACTTTGCAAGGGAGTGGTGCTGAGATCCATCAGCCCCAGGGGGTCAGGCAAG
GGGTCAGAGGCACCCTGGGAGGTTTGTACTGGGCTGAAATCCAGTCCCCCTACTTTGGCTGGGGGCGTGA
GCCTCCAGGATTCAGGGGTTCTGGGGAGGACAGATTTGCTCGGGGTGGAGGAGATGGGCAGCGTTTCCTT
AATGGGTGTCTTAAAAGGTCCTCCCACTTCCTGGGAGTAGCTGAGCTGGGAGGCAGGGTCAGAGGAGTCT
GCTGGGAACGGGAGCTCTGCGGCCCAGCGGGAAGTACTGGGCCCCTCTGAGAAGAGCAGCTCCGGCTCAT
CCACACAGGGAGGCAGTAGATGCTGTTTCCTCCGAGACCGGCTCCTCTCCCTCCTCTCCCTGTGTTGAAT
CACAAGCATTTCCGAGACACACCGGGTTGGGGACCTAAGCCCACTGTAGGACTTCTTGGGTCTTGGGGTG

GGAGATTGGGACGAATCCTCCCAGGAGTGAGATGATTCCTCTTTGAAAGATGGGGCCGGGGAGGGCCACT
CTTCCAAGGGAGGGCTCTCCACTTTGATGGGTCTCGCTAAGTGTGGCATTTCCTCCCCAGGCTGGATTTC
TTCCTCCTTGATAGTCTGAACTGGAAGCAAAGGAGAAAACCCTTCTCCAAACAGGAGTTTCTCCTCTTTC
CCTGGTCCTGCAGAAGAAAGAGGAGCTATCCCGTCCTCAGCTAGCAGCACCTTGGGGGCAATGCGGACTC
GCTTGCTATGGCGGGCAAGCTCTGAGCTCATGAGGGAAGCCGCCAGGGGCAATG

>DEX0356_025.nt.2
TTCGTCTCGCTTTGTTCTGGTTCTTCCTCGTGTCGGTCGTTATTGTTCTTCTTCCGCTTCGTTCTCATTC
GTTGGTCCTTTCCCTTGCTCTCTTCCTCCTTTTCCTCCTCCCTCTTCGTGGTTCTCTTCCGGCGCTGTGG
TCTTGTTCGTTTCTCTCGGTCTGTCCTGGCTTCCGTCTTGTTGGCTCTTCTCCTTCTGTCTTCTTGCGTC
CGGTTTCCTGTTGCTTGTCTGTCTTTCTCGCTTCTGCTCGTCATCTGTTTCTCCTTGTTCTTGCTTTTTC
TCTCTCCGGTCTCGCCTTCTTTCCTTGTCTCCTCCTCCCCTTTTCTTCTCTTTGCGTGTGCTTGTCTCGC
TCGCTCTGTGTTCTTCTGTCTCTGCTTTTGTCGTGTCCGCTTGTCTCTTGTCTTCTTCGGCTTGCTTTTT
CTCTTCTCTCCTCTCCGTTCGCTGCTGTTTTCTGTTCTCCGTGCCTCTGTCCCTTTTGTCTTCTTCGTCT
TCTTCGCTTCCTTCCGTTCCCTCCGCTCCTCTTCCTCCGTTCCTCTCCTCTCGTCGTTCCTCCCGCTCTC
CCCTTTCGCTGCTGCTCTGGCTTCCCTCGCTGGCCGTCCTCCCTCTTCGCCTCCCTCTCCTTCCCTCTGT
CGTTTCTCGTTGTTGTTCTTGTGTGTTGTTGTGTGTGCTCGTGTTGTTCTGGTTCTTGGTTGGTGGTTGT
GTGGTGTGTGCGCTGTGTGTGTTGTTTGTTGTGTTTGTGCGCTCGTGGTGTACTGCTGAGAAGTCACACC
ACCAGCGTACGTCGTTTAACCGGCTATCGTGGGCGCATCCCCCGAGGGTCTCCGAGCCGGGCGGTCCGGC
GGCTGTTCCCGCCTTCTTTTTTTTGCTCCTTGGGCCCTGTCCAAACGGTCCCGTACTTGGCCCTCGAGGG
AACTTTGGCGCCCCCTTTTTTTTTTTGCATGTCCACCTTCGCTTTTATTGAGTAGTTAGTGTTCTCAAGC
TGGCTCACACCCATGCGGTCAGCTGGGGCCTAGGGTGGCTCTTTGCAAAGCTGAGGGGCAAGCTAAGGAA
GCCAGGCAGGTCAGGGGCCCTTTCGGCCTTCTCAAGCCTCCACCTGAGTTCTCGTCAATGCCAGTCTCCC
TGGTATGATTGGGGACATTATCAGAGAAACATCTAATAGCGCACATCTGGGCACCCACACTCTGCTTCAG
TTGCATCCATCCTCCCACCCCAAATTCAACTCCTGACCCAATACAAAAGACTTTTTTAACCAGGATTTCT
TCTTGCAGGAAAGCTGACTTGGAAACACGGGGAGGTGGCAGGGAGGGGCAAAAAGGACTCTGGCAAGCAG
ATCCACTTGTCTGGGTCCCTGCAGTGAAGAACCCAAGATCCAGGTACCTCAGCCTGGAAGAAACCGTGCA
CTGCAGGTCTTCCCTTCTATCAAAACTAAAAGCAAGGAACAGAAGTCAATTGAAACCCCAGGTCATCCTA
CCCACCTTCTGGCAGTCTCTGGATAATGATCTAAGGAAATAAGCTTACATTTATAATTAGAGGATAATTT
GGAGAATTTATACTATTTACACGGACCACCCTGCAAAGATCAGGGAAGGTAAGAGACTGCTGGGCAGAGA
ATGGAAACAGGCTGGGGGGTTCCTAATCTCTTTTCACCATTGCCTTTGTTGTTCCCACCCTTCAGCTCCT
GGCAGGGACAGCAGAGGAATCAGATACTCTTGGGGAACACAAGGTCCCAGCAGTGGCTAGGGTGTGACTG
CTACTTTTGCATAATCAGGCAGCAGGGAGCTATGAGGAGCAGAACAGTCCCTGCCTGCTGTCCTCACTCA
GAGGCTTGGGGTGCACTGAGCCTTGGAGTGCCCGGGATGGTGGACAGCTTGAGCACAGGGGCAAGGGCAG
GGCTCTACTGTAGCTCAGGAATAAACTGGGACCAGTTGATGTTGTCAGGGCCCAGTGGGTCCTCGTCCAG
GCCAGGAAAGCTGATGTCCAGCAGGATCTTGCTGAGGCTGTCATTCATTGTGTCCAGGACCAGGCCTTCT
CAACCCCAGCCTCGYWCCTAGCG

>DEX0356_026.nt.1
TTCGAGTCGGTACCCGCAAGGTGAATATGAGACCTGCAATAGCGATAATCTGAACCATGCGGTTTTGGCA
GTGGGATATGGAATCCAGAAGGGAAACAAGCACTGGATAATTAAAAACAGCTGGGGAGAAAACTGGGGAA
ACAAAGGATATATCCTCATGGCTCGAAATAAGAACAACGCCTGTGGCATTGCCAACCTGGCCAGCTTCCC
CAAGATGTGACTCCAGCCAGCCAAATCCATCCTGCTCTTCCATTTCTTCCACGATGGTGCAGTGTAACGA
TGCACTTTGGAAGGGAGTTGGTGTGCTATTTTTGAAGCAGATGTGGTGATACTGAGATTGTCTGTTCAGT
TTCCCCATTTGTTTGTGCTTCAAATGATCCTTCCTACTTTGCTTCTCTCCACCCATGACCTTTTTCACTG
TGGCCATCAGGACTTTCCCTGACAGCTGTGTACTCTTAGGCTAAGAGATGTGACTACAGCCTGCCCCTGA
CTGTGTTGTCCCAGGGCTGATGCTGTACAGGTACAGGCTGGAGATTTTCACATAGGTTAGATTCTCATTC
ACGGGACTAGTTAGCTTTAAGCACCCTAGAGGACTAGGGTAATCTGACTTCTCACTTCCTAAGTTCCCTT
CTATATCCTCAAGGTAGAAATGTCTATGTTTTCTACTCCAATTCATAAATCTATTCATAAGTCTTTGGTA
CAAGTTTACATGATAAAAAGAAATGTGATTTGTCTTCCCTTCTTTGCACTTTTGAAATAAAGTATTTATC
TCCTGTCTACAGTTTAATAAATAGCATCTAGTACACATTCATTTTGTGTTGGATACTGTGTTAGGTGCTG
GAGGAAAAAAGATGAATAGAACATCTTCTATGTACTTGATGCGCTCACAGTCTGGTTGTAGAGACTGTCA
CATAAACATTTCATCCCAATTC

>DEX0356_027.nt.1
AGAAGAGGACAGGAAAGAAACTTCAGTCTGAGGGTTGAGGACACGTCCCGGAGCAGAGGAGGACCAAGCT
CCCAAGCCCCCTGTCACCTTGACACGCAGACGGTTCCCCAGGACTGCGAAGAGGGGGGCAACCAGCGCCC
CGCCCCGCCCCGCTGCGCCGCCCGCCATTGGCCGGTCCGAAAGACGCGGGGCGTTGTCGGTTTGAATGAG
AGCGCCTGGAGCCGGGGCGCAGGCGCAGAGGAGGGCGGTGTTGAGACCGGCGGAGCGGCGGGACCCCTAG
GTGGCGGAGGGACGCTCCGGGAAAGCGAGGGGCGCTACGAGCTCTGGCCCACGTGACCTGCCGGGGGCGG

GAGCAGGGGGCGCGCCGGCCTCCTGCGGTGCCCCTGCCTTGGGGAGGGGCCGTGACCACCCGTCTGTCGC
CCGAGGCGGCCGCCGCTGCACCTTCACCGCGTACCCGGGACCCGCCCGCCCGCGGGAGCTTATAATATGA
CTCGATGGAGGAAAATTTGATAAGCATGAGAGAAGACCATTCTTTTCATGTTCGTTACAGAATGGAAGCT
TCTTGCCTAGAGCTGGCCTTGGAAGGGGAACGTCTATGTAAATCAGGAGACTGCCGCGCTGGCGTGTCAT
TCTTTGAAGCTGCAGTTCAAGTTGGAACTGAAGACCTAAAAACACTTAGCGCTATTTACAGCCAGTTGGG
CAATGCTTATTTCTATTTGCATGATTATGCCAAAGCATTAGAATATCACCATCATGATTTAACCCTTGCA
AGGACTATTGGAGACCAGCTGGGGGAAGCGAAAGCTAGTGGTAATCTGGGAAACACCTTAAAAGTTCTTG
GGAATTTTGACGAAGCCATAGTTTGTTGTCAGCGACACCTAGATATTTCCAGAGAGCTTAATGACAAGGT
GGGAGAAGCAAGAGCACTTTACAATCTTGGGAATGTGTATCATGCCAAAGGGAAAAGTTTTGGTTGCCCT
GGTCCCCAGGATGTAGGAGAATTTCCAGAAGAAGTGAGAGATGCTCTGCAGGCAGCCGTGGATTTTTATG
AGGAAAACCTATCATTAGTGACTGCTTTGGGTGACCGAGCGGCACAAGGACGTGCCTTTGGAAATCTTGG
AAACACACATTACCTCCTTGGCAACTTCAGGGATGCAGTTATAGCTCATGAGCAGCGTCTCCTTATTGCA
AAAGAATTTGGAGATAAAGCAGCTGAAAGAAGAGCATATAGCAACCTTGGAAATGCATATATATTTCTTG
GTGAATTTGAAACTGCCTCGGAATACTACAAGAAGACACTACTGTTGGCCCGACAGCTTAAAGACCGAGC
TGTAGAAGCACAGTCTTGTTACAGTCTTGGAAATACATATACTTTACTTCAAGACTATGAAAAGGCCATT
GATTATCATCTGAAGCACTTAGCAATTGCTCAAGAGCTGAATGATAGAATTGGTGAAGGAAGAGCATGTT
GGAGCTTAGGAAATGCATACACAGCACTAGGAAATCATGATCAAGCAATGCATTTTGCTGAAAAGCACTT
GGAAATTTCAAGAGAGGTTGGGGATAAAAGTGGTGAACTAACAGCACGACTTAATCTCTCAGACCTTCAA
ATGGTTCTTGGTCTGAGCTACAGCACAAATAACTCCATAATGTCTGAAAATACTGAAATTGATAGCAGTT
TGAATGGTGTACGCCCCAAGTTGGGACGCCGGCATAGTATGGAAAATATGGAACTTATGAAGTTAACACC
AGAAAAGGTACAGAACTGGAACAGTGAAATTCTTGCTAAGCAAAAACCTCTTATTGCCAAACCTTCTGCA
AAGCTACTCTTTGTCAACAGACTGAAGGGGAAAAAATACAAAACGAATTCCTCCACTAAAGTTCTCCAAG
ATGCCAGTAATTCTATTGACCACCGAATTCCAAATTCTCAGAGGAAAATCAGTGCAGATACTATTGGAGA
TGAAGGGTTCTTTGACTTATTAAGCCGATTTCAAAGCAATAGGATGGATGATCAGAGATGTTGCTTACAA
GAAAAGAACTGCCATACAGCTTCAACAACAACTTCTTCCACTCCCCCTAAAATGATGCTAAAAACATCAT
CTGTTCCTGTGGTATCCCCCAACACGGATGAGTTTTTAGATCTTCTTGCCAGCTCACAGAGTCGCCGTCT
GGATGACCAGAGGGCTAGTTTCAGTAATTGCCAGGGCTTCGTCTAACACAAAACAGCCAGTCGGTACTT
AGCCACCTGATGACTAATGACAACAAAGAGGCTGATGAAGATTTCTTTGACATCCTTGTAAAATGTCAAG
GATCCAGATTAGATGATCAAAGATGTGCTCCACCACCTGCTACCACAAAGGGTCCGACAGTACCAGATGA
AGACTTTTTCAGCCTTATTTTACGGTCCCAGGGAAAGAGAATGGATGAACAGAGAGTTCTTTTACAAAGA
GATCAAAACAGAGACACTGACTTTGGGCTAAAGGACTTTTTGCAAAATAATGCTTTGTTGGAGTTTAAAA
ATTCAGGGAAAAAATCGGCAGACCATTAGTTACTATGGATTTATTTTTTTCCTTTCAAACACGGTAAGG
AAACAATCTATTACTTTTTTCCTTAAAAGGAGAATTTATAGCACTGTAATACAGCTTAAAATATTTTTAG
AATGATGTAAATAGTTAACCTTCAGTAGTCTATTAAGGCATTAATACTTCTCTGGACATGCGCGTTTGAG
GGTGGAGGGGTCCTGTAAGGTGCTTCATCGTCTGTGATTACTGCTTGGGATGTGTTCTTTGGCAGCTTGT
GAGATTACTTTACCTAGTGTTTATAAAGTAGGAAGTTAAGTGAATCATAGATTAGAATTTAATACTCTTA
TGGAAATAATTTTTTAACATCTTAATTGACAATGGCGTTTTTTTATACATAACCATGGATGTAGTGGGAA
ACAATGTTGTTTGGTAAAAATAATGTACTTGATCAATGTAAAAAAGTATATAAAATAGT

>DEX0356_027.nt.2
AGAAGAGGACAGGAAAGAAACTTCAGTCTGAGGGTTGAGGACACGTCCCGGAGCAGAGGAGGACCAAGCT
CCCAAGCCCCCTGTCACCTTGACACGCAGACGGTTCCCCAGGACTGCGAAGAGGGGGGCAACCAGCGCCC
CGCCCCGCCCCGCTGCGCCGCCCGCCATTGGCCGGTCCGAAAGACGCGGGGCGTTGTCGGTTTGAATGAG
AGCGCCTGGAGCCGGGGCGCAGGCGCAGAGGAGGGCGGTGTTGAGACCGGCGGAGCGGCGGGACCCCTAG
GTGGCGGAGGGACGCTCCGGGAAAGCGAGGGGCGCTACGAGCTCTGGCCCACGTGACCTGCCGGGGGCGG
GAGCAGGGGGCGCGCCGGCCTCCTGCGGTGCCCCTGCCTTGGGGAGGGGCCGTGACCACCCGTCTGTCGC
CCGAGGCGGCCGCCGCTGCACCTTCACCGCGTACCCGGGACCCGCCCGCCCGCGGGAGAACCCACAGATC
TGGAGGACCTACTATAATTTAATTTAATTCTAAGCTTATAATATGACTCGATGGAGGAAAATTTGATAAG
CATGAGAGAAGACCATTCTTTTCATGTTCGTTACAGAATGGAAGCTTCTTGCCTAGAGCTGGCCTTGGAA
GGGGAACGTCTATGTAAATCAGGAGACTGCCGCGCTGGCGTGTCATTCTTTGAAGCTGCAGTTCAAGTTG
GAACTGAAGACCTAAAAACACTTAGCGCTATTTACAGCCAGTTGGGCAATGCTTATTTCTATTTGCATGA
TTATGCCAAAGCATTAGAATATCACCATCATGATTTAACCCTTGCAAGGACTATTGGAGACCAGCTGGGG
GAAGCGAAAGCTAGTGGTAATCTGGGAAACACCTTAAAAGTTCTTGGGAATTTTGACGAAGCCATAGTTT
GTTGTCAGCGACACCTAGATATTTCCAGAGAGCTTAATGACAAGGTGGGAGAAGCAAGAGCACTTTACAA
TCTTGGGAATGTGTATCATGCCAAAGGGAAAAGTTTTGGTTGCCCTGGTCCCCAGGATGTAGGAGAATTT
CCAGAAGAAGTGAGAGATGCTCTGCAGGCAGCCGTGGATTTTTATGAGGAAAACCTATCATTAGTGACTG
CTTTGGGTGACCGAGCGGCACAAGGACGTGCCTTTGGAAATCTTGGAAACACACATTACCTCCTTGGCAA
CTTCAGGGATGCAGTTATAGCTCATGAGCAGCGTCTCCTTATTGCAAAAGAATTTGGAGATAAAGCAGCT
GAAAGAAGAGCATATAGCAACCTTGGAAATGCATATATATTTCTTGGTGAATTTGAAACTGCCTCGGAAT
ACTACAAGAAGACACTACTGTTGGCCCGACAGCTTAAAGACCGAGCTGTAGAAGCACAGTCTTGTTACAG

TCTTGGAAATACATATACTTTACTTCAAGACTATGAAAAGGCCATTGATTATCATCTGAAGCACTTAGCA
ATTGCTCAAGAGCTGAATGATAGAATTGGTGAAGGAAGAGCATGTTGGAGCTTAGGAAATGCATACACAG
CACTAGGAAATCATGATCAAGCAATGCATTTTGCTGAAAAGCACTTGGAAATTTCAAGAGAGGTTGGGGA
TAAAAGTGGTGAACTAACAGCACGACTTAATCTCTCAGACCTTCAAATGGTTCTTGGTCTGAGCTACAGC
ACAAATAACTCCATAATGTCTGAAAATACTGAAATTGATAGCAGTTTGAATGGTGTACGCCCCAAGTTGG
GACGCCGGCATAGTATGGAAAATATGGAACTTATGAAGTTAACACCAGAAAAGGTACAGAACTGGAACAG
TGAAATTCTTGCTAAGCAAAAACCTCTTATTGCCAAACCTTCTGCAAAGCTACTCTTTGTCAACAGACTG
AAGGGGAAAAAATACAAAACGAATTCCTCCACTAAAGTTCTCCAAGATGCCAGTAATTCTATTGACCACC
GAATTCCAAATTCTCAGAGGAAAATCAGTGCAGATACTATTGGAGATGAAGGGTTCTTTGACTTATTAAG
CCGATTTCAAAGCAATAGGATGGATGATCAGAGATGTTGCTTACAAGAAAAGAACTGCCATACAGCTTCA
ACAACAACTTCTTCCACTCCCCCTAAAATGATGCTAAAAACATCATCTGTTCCTGTGGTATCCCCCAACA
CGGATGAGTTTTTAGATCTTCTTGCCAGCTCACAGAGTCGCCGTCTGGATGACCAGAGGGCTAGTTTCAG
TAATTTGCCAGGGCTTCGTCTAACACAAAACAGCCAGTCGGTACTTAGCCACCTGATGACTAATGACAAC
AAAGAGGCTGATGAAGATTTCTTTGACATCCTTGTAAAATGTCAAGGATCCAGATTAGATGATCAAAGAT
GTGCTCCACCACCTGCTACCACAAAGGGTCCGACAGTACCAGATGAAGACTTTTTCAGCCTTATTTTACG
GTCCCAGGGAAAGAGAATGGATGAACAGAGAGTTCTTTTACAAAGAGATCAAAACAGAGACACTGACTTT
GGGCTAAAGGACTTTTTGCAAAATAATGCTTTGTTGGAGTTTAAAAATTCAGGGAAAAAATCGGCAGACC
ATTAGTTACTATGGATTTATTTTTTTTCCTTTCAAACACGGTAAGGAAACAATCTATTACTTTTTTCCTT
AAAAGGAGAATTTATAGCACTGTAATACAGCTTAAAATATTTTTAGAATGATGTAAATAGTTAACCTTCA
GTAGTCTATTAAGGCATTAATACTTCTCTGGACATGCGCGTTTGAGGGTGGAGGGGTCCTGTAAGGTGCT
TCATCGTCTGTGATTACTGCTTGGGATGTGTTCTTTGGCAGCTTGTGAGATTACTTTACCTAGTGTTTAT
AAAGTAGGAAGTTAAGTGAATCATAGATTAGAATTTAATACTCTTATGGAAATAATTTTTTAACATCTTA
ATTGACAATGGCGTTTTTTTATACATAACCATGGATGTAGTGGGAAACAATGTTGTTTGGTAAAAATAAT
GTACTTGATCAATGTAAAAAAGTATATAAAATAGT

>DEX0356_028.nt.1
TGCAAGTTGGGGCCCAAGGCAGGGGAAGGATCCGGCAAAAAAGGCACACGGTTCAGGCCTGTCCACTCCA
GCCTTGGGAGGAGGTGAGCAGGTCACCATGGCTCGGAGCCTCTCCATCATGTTCGTCCCCAACAGCTGGC
TCAATGGCTACCACAAGGTGGCCATGAACGGTAAGGCCCGCAAAGAGGCGGTGCAGACTGCGGCTAAGGA
ACTCCTCAAGTTCGTGAACCGGAGTCCCTCTCCTTTCCATGCTGTGGCTGAATGCCGCAACCGCCTTCTC
CAGGCTGGCTTCAGTGAACTCAAGGAGACTGAGAAATGGAATATTAAGCCCGAGAGCAAGTACTTCATGA
CCAGGAACTCCTCCACCATCATAGCTTTTGCTGTAGGGGGCCAGTACGTTCCTGGCAATGGCTTCAGCCT
CATCGGGGCCCACACGGACAGCCCCTGCCTCCGGGTGAAACGTCGGTCTCGCCGCAGCCAGGTGGGCTTC
CAGCAAGTCGGTGTGGAGACCTATGGTGGTGGGATCTGGAGCACCTGGTTTGACCGTGACCTGACTCTGG
CTGGACGCGTCATTGTCAAGTGCCCTACCTCAGGTCGGCTGGAGCAGCAGCTGGTGCACGTGGAGCGGCC
CATTCTTCGCATCCCACACCTGGCCATCCATCTGCAGCGAAATATCAACGAGAACTTTGGGCCCAACACA
GAGATGCATCTAGTCCCCATTCTTGCCACAGCCATCCAGGAGGAGCTGGAGAAGGGGACTCCTGAGCCAG
GGCCTCTCAATGCTGTGGATGAGCGGCACCATTCGGTCCTCATGTCCCTGCTCTGTGCCCATCTGGGGCT
GAGCCCCAAGGACATAGTGGAGATGGAGCTCTGCCTTGCAGACACCCAGCCTGCGGTCTTGGGTGGTGCC
TATGATGAGTTCATCTTTGCTCCTCGGCTGGACAATCTGCACAGCTGCTTCTGTGCCCTGCAGGCCTTGA
TAGATTCCTGTGCAGGCCCTGGCTCCCTGGCCACAGAGCCTCACGTGCGCATGGTCACACTCTATGACAA
CGAAGAGGTGGGGTCTGAGAGTGCACAGGGAGCACAGTCACTGCTGACAGAGCTGGTGCTGCGGCGGATC
TCAGCCTCGTGCCAGCACCCGACAGCCTTCGAGGAAGCCATACCCAAGTCCTTCATGATCAGCGCAGACA
TGGCCCATGCTGTGCATCCCAACTACCTGGACAAGCATGAGGAGAACCACCGGCCTTTATTCCACAAGGG
CCCCGTGATCAAGGTGAACAGCAAGCAACGCTATGCTTCAAACGCGGTGTCAGAGGCCCTGATCCGAGAG
GTGGCCAACAAAGTCAAGGTCCCCCTGCAGGATCTCATGGTCCGGAATGACACCCCCTGTGGAACCACCA
TTGGACCTATCTTGGCTTCTCGGCTGGGGCTGCGGGTGCTGGATTTAGGCAGCCCCCAACTGGCCATGCA
CTCTATCCGGGAGATGGCCTGCACCACAGGAGTCCTCCAGACCCTCACCCTCTTCAAGGGCTTCTTTGAG
CTGTTCCCTTCTCTAAGCCATAATCTCTTAGTGGATTGAGCCCTCTTGGAAAGACTTCTCTGCCATCCCT
TTGCACCTGAGAGGGGAAGTTCTCAGCTGAGCTGAAGCTGGATTATTAAAGTGGATTGTCACTCAGA

>DEX0356_028.nt.2
GAGGCCAAACATGGGGCGGGGCGGCGCGGCCGGGGAAGCGTGATGAAGGCCTACGAGTGCGGCGCGGCCT
GAAGGGGCACGCGGGGGACCTGCAAAGCTAGTGAGGGGCGGGGCAGGCGGCGCGGTGGGGGCGGGCCGAG
CCCGGAGGCCAGATGAGCGGACACAGCCCCACGCGCGGGGCCATGCAGGTGGCCATGAACGGTAAGGCCC
GCAAAGAGGCGGTGCAGACTGCGGCTAAGGAACTCCTCAAGTTCGTGAACCGGAGTCCCTCTCCTTTCCA
TGCTGTGGCTGAATGCCGCAACCGCCTTCTCCAGGCTGGCTTCAGTGAACTCAAGGAGACTGAGAAATGG
AATATTAAGCCCGAGAGCAAGTACTTCATGACCAGGAACTCCTCCACCATCATAGCTTTTGCTGTAGGGG
GCCAGTACGTTCCTGGCAATGGCTTCAGCCTCATCGGGGCCCACACGGACAGCCCCTGCCTCCGGGTGAA
ACGTCGGTCTCGCCGCAGCCAGGTGGGCTTCCAGCAAGTCGGTGTGGAGACCTATGGTGGTGGGATCTGG

AGCACCTGGTTTGACCGTGACCTGACTCTGGCTGGACGCGTCATTGTCAAGTGCCCTACCTCAGGTCGGC
TGGAGCAGCAGCTGGTGCACGTGGAGCGGCCCATTCTTCGCATCCCACACCTGGCCATCCATCTGCAGCG
AAATATCAACGAGAACTTTGGGCCCAACACAGAGATGCATCTAGTCCCCATTCTTGCCACAGCCATCCAG
GAGGAGCTGGAGAAGGGGACTCCTGAGCCAGGGCCTCTCAATGCTGTGGATGAGCGGCACCATTCGGTCC
TCATGTCCCTGCTCTGTGCCCATCTGGGGCTGAGCCCCAAGGACATAGTGGAGATGGAGCTCTGCCTTGC
AGACACCCAGCCTGCGGTCTTGGGTGGTGCCTATGATGAGTTCATCTTTGCTCCTCGGCTGGACAATCTG
CACAGCTGCTTCTGTGCCCTGCAGGCCTTGATAGATTCCTGTGCAGGCCCTGGCTCCCTGGCCACAGAGC
CTCACGTGCGCATGGTCACACTCTATGACAACGAAGAGGTGGGGTCTGAGAGTGCACAGGGAGCACAGTC
ACTGCTGACAGAGCTGGTGCTGCGGCGGATCTCAGCCTCGTGCCAGCACCCGACAGCCTTCGAGGAAGCC
ATACCCAAGTCCTTCATGATCAGCGCAGACATGGCCCATGCTGTGCATCCCAACTACCTGGACAAGCATG
AGGAGAACCACCGGCCTTTATTCCACAAGGGCCCCGTGATCAAGGTGGCAAAGGCCCCAACGTTGTAGGC
CCTACGGGCTACTACAACCCTTCGGTGACGCCATAAAAACTCTTCACCAAAGAGCCCTTAAACCCGCAAA
TCTACCATCAACCTTTAAATCAACGGCCGACTTAGCTTTCAACATCGTCTTATTATTGAAACCCCCCTCC
ATAGCGAACCCCGGGTCAACTACACTAAGGTCTATTTATTGAGGCCCTTACGAGCGGGCACCCACCCCTG
CCAGGGGGGAACCGCTGCCGTGGCCGGCGACGCCCCACACTCTTCGCCCATTTTGAGTGCCACACGCGCC
GCGCAAAAGCACAAACAACAAGGGGCACGGCGCCGCATGA

>DEX0356_029.nt.1
CTCTACAGGATCCCCGGTGTCTCGCCACAMAGATTMCATCCTCATGCATCTAATCACAGACGATGAAGCA
CCTTACAGGACCCCTCCACCCTCAAACGCGCATGTCCAGAGAAGTATTAATGCCTTAATAGACTACTGAA
GGTTAACTATTTACATCATTCTAAAAATATTTTAAGCTGTATTACAGTGCTATAAATTCTCCTTTTAAGG
AAAAAAGTAATAGATTGTTTCCTTACCGTGTTTGAAAGGAAAAAAAATAAATCCATAGTAACTAATGGTC
TGCCGATTTTTTCCCTGAATTTTTAAACTCCAACAAAGCATTATTTTGCAAAAAGTCCTTTAGCCCAAAG
TCAGTGTCTCTGTTTTGATCTCTTTGTAAAAGAACTCTCTGTTCATCCATTCTCTTTCCCTGGGACCGTA
AAATAAGGCTGAAAAAGTCTTCATCTGGTACTGTCGGACCCTTTGTGGTAGCAGGTGGTGGAGCACATCT
TTGATCATCTAATCTGGATCCCTAGAAGTTTATGAGAGAGATTAAAAGACATACTGTGGGAAAGAAAGCA
CCAAGCTATGTTAAATCTCTCAGTCACAACAAGAAAACGCAGTTGTAAACTTTTATTTGGGGCTTCTCAG
AAATTAATCTTTGTTTTTATATTGTTCTCTTTAGTTCCCAGCCTGATTTCTGAAAGAAATAATCTTTAAA
CCTGAAACTGCATAAAATTTTTCCTTTAAAATGGCTTATATAGGTATAGCTGCTTCATTGGTTTAGTACG
TGCATTATGTTAAGTCATACTGTTTTAATTATAATTGGTCACAAGGATAAAAGACAAGGTCAACTGTATC
TAAGTATGACAGTTCAAGGTTATAATCTAACAGGCATCAACTGTGTAACTTAACAAATCTTTAGAAATAT
GTACAAACCCCATGAGGAACTACCTTAAAAAAATCACGTCTTAGGCTAGGGGACAGCAAACTGTAACCCA
GTGCCGGCTTCTGTAAATGAAGTCATTTTCATTTTCCCATCAATACATATGTTTATATACTATCTATATG
GCTGCTTCCCTACTAAATGGCAGAGTTGGGCAGTTGCAATAGATACCATATGGGCGGCAAAGCCTAAAAT
ATTTTCTATCTGGCCTTTTACAGAAAAAGTTTGCCTGCCTCTGTCTTAGGCTGTTAAATCTGTTAAGAAA
TAGCCCAGAATGAA

>DEX0356_030.nt.1
CACTAGTACGGCGCAGTGTGCGGAATTCGGTTTCGAGCGGCGCCCGGGAGTACTGTGAATAATGATGAAG
GAGAGTGGAGTGGCCCACCACCTGAATGCAGAGGAAAATCTCTAACTTCCAAGGTCCCACCAACAGTTCA
GAAACCTACCACAGTAAATGTTCCAACTACAGAAGTCTCACCAACTTCTCAGAAAACCACCACAAAAACC
ACCACACCAAATGCTCAAGCAACACGGAGTACCTCGGCCGCGACCACGCTAAGCC

>DEX0356_031.nt.1
GGCTTAGCGTGGTCGCGGCCGAGGTACAGGGTGGGAGATGGGGGAGGGGCTGGGGGCTGAAGGGAGCAAG
AAAGCTGTAGCTGTGTGGGGCTGGCAGGATGTTGAAGACCGCCTCTGCTGCTCTCACATGGGGACTGGGC
CCAGATCCTGCTTGGTCACACCCAGCCCAGAAGACGGGTCCTCCAGTTCCAGTGACTCACTGCAGTTATG
GTGCCCGCGGCTTCTGGTGTTGGGGACCTCCGTGCCGTTGGGGGGAGGAGAAGCAGGAGCTGTCGGGAAG
ATCAGAAGCCAGTCATGGATGACCAGCGCGACCTTATCTCCAACAATGAGCAACTGCCCATGCTGGGCCG
GCGCCCTGGGGCCCCGGAGAGCAAGTGCAGCCGCGGAGCCCTGTACACAGGCTTTTCCATCCTGGTGACT
CTGCTCCTCGCTGGCCAGGCCACCACCGCCTACTTCCTGTACCAGCAGCAGGGCCGGCTGGACAAACTGA
CAGTCACCTCCCAGAACCTGCAGCTGGAGAACCTGCGCATGAAGCTTCCCAAGCCTCCCAAGCCTGTGAG
CAAGATGCGCATGGCCACCCCGCTGCTGATGCAGGCGCTGCCCATGGGAGCCCTGCCCCAGGGGCCCATG
CAGAATGCCACCAAGTATGGCAACATGACAGAGGACCATGTGATGCACCTGCTCCAGAATGCTGACCCCC
TGAAGGTGTACCCGCCACTGAAGGGGAGCTTCCCGGAGAACCTGAGACACCTTAAGAACACCATGGAGAC
CATAGACTGGAAGGTCTTTGAGAGCTGGATGCACCATTGGCTCCTGTTTGAAATGAGCAGGCACTCCTTG
GAGCAAAAGCCCACTGACGCTCCACCGAAAGTACTGACCAAGTGCCAGGAAGAGGTCAGCCACATCCCTG
CTGTCCACCCGGGTTCATTCAGGCCCAAGTGCGACGAGAACGGCAACTATCTGCCACTCCAGTGCTATGG
GAGCATCGGCTACTGCTGGTGTGTCTTCCCCAACGGCACGGAGGTCCCCAACACCAGAAGCCGCGGGCAC
CATAACTGCAGTGGTAAGCAGTGGCACTGTGCCAGTGTCAGAGGACCAGGAAGGACTAGGAAGGTTGAGG

GGCAAGAGGTCCCCTCTGAAGCACATGGGACCAGGACACCCAGGATGGCAGCTCCTGGGGGCAGTGACGT
AGTCATGCGTCCAGCCCTTTATCCATCCACCCACCTGTGCATTCCTATTTGTCCATTTATCATCTCTCCT
CTTACATGCACTCATATTTGTTCACTTATCCGTCAGTCCTTTCATCTGTGAATTTCATCCATCCACGCGC
CATCCTCTAGCATCCAGGAGTCCTACAGACCCACCCATCTCATTTCATCACCCCCTTCTCACTCGAGCCC
CCATTACACCTCTTTGTTGCTGCTTGCAGCTGTCCTTCCCTGGGTACCTGCTTTCCCAGGCACTAAGCCT
GTGGCTAGGTGGGAAGCACTGCCCTCAGGCATCTTGGGTCAGGTAGGCTGCACTTCAAGTGACAAACGGA
CTTGCTGCTCCTTTGCAGAGTCACTGGAACTGGAGGACCCGTCTTCTGGGCTGGGTGTGACCAAGCAGGA
TCTGGGCCCAGTCCCCATGTGAGAGCAGCAGAGGCGGTCTTCAACATCCTGCCAGCCCCACACAGCTACA
GCTTTCTTGCTCCCTTCAGCCCCCAGCCCCTCCCCCATCTCCCACCCTGTACCTCATCCCATGAGACCCT
GGTGCCTGGCTCTTTCGTCACCCTTGGACAAGACAAACCAAGTCGGAACAGCAGATAACAATGCAGCAAG
GCCCTGCTGCCCAATCTCCATCTGTCAACAGGGGCGTGAGGTCCCAGGAAGTGGCCAAAAGCTAGACAGA
TCCCCGTTCCTGACATCACAGCAGCCTCCAACACAAGGCTCCAAGACCTAGGCTCATGGACGAGATGGGA
AGGCACAGGGAGAAGGGATAACCCTACACCCAGACCCCAGGCTGGACATGCTGACTGTCCTCTCCCCTCC
AGCCTTTGGCCTTGGCTTTTCTAGCCTATTTACCTGCAGGCTGAGCCACTCTCTTCCCTTTCCCCAGCAT
CACTCCCCAAGGAAGAGCCAATGTTTTCCACCCATAATCCTTTCTGCCGACCCCTAGTTCCCTCTGCTCA
GCCAAGCTTGTTATCAGCTTTCAGGGCCATGGTTCACATTAGAATAAAAGGTAGTAATAGAAAAAAAAAA
AAAACCCGGGCCTTATGGCGGGGGGGAAAGGGGACAGGAAGGGCATGGCTAGGGTTCGGCCGGTCTTGAC
AGGGGATTGCGGGCCA

>DEX0356_031.nt.2
ATTCTGCATGGGCCCTGGGGCAGGGCTSSCATGGGCAGCGCTGCATCAGAATGCTGACCCCCTGAAGGTG
TACCCGCCACTGAAGGGGAGCTTCCCGGAGAACCTGAGACACCTTAAGAACACCATGGAGACCATAGACT
GGAAGGTCTTTGAGAGCTGGATGCACCATTGGCTCCTGTTTGAAATGAGCAGGCACTCCTTGGAGCAAAA
GCCCACTGACGCTCCACCGAAAGAGTCACTGGAACTGGAGGACCCGTCTTCTGGGCTGGGTGTGACCAAG
CAGGATCTGGGCCCAGTCCCCATGTGAGAGCAGCAGAGGCGGTCTTCAACATCCTGCCAGCCCCACACAG
CTACAGCTTTCTTGCTCCCTTCAGCCCCCAGCCCCTCCCCCATCTCCCACCCTGTACCTCATCCCATGAG
ACCCTGGTGCCTGGCTCTTTCGTCACCCTTGGACAAGACAAACCAAGTCGGAACAGCAGATAACAATGCA
GCAAGGCCCTGCTGCCCAATCTCCATCTGTCAACAGGGGCGTGAGGTCCCAGGAAGTGGCCAAAAGCTAG
ACAGATCCCCGTTCCTGACATCACAGCAGCCTCCAACACAAGGCTCCAAGACCTAGGCTCATGGACGAGA
TGGGAAGGCACAGGGAGAAGGGATAACCCTACACCCAGACCCCAGGCTGGACATGCTGACTGTCCTCTCC
CCTCCAGCCTTTGGCCTTGGCTTTTCTAGCCTATTTACCTGCAGGCTGAGCCACTCTCTTCCCTTTCCCC
AGCATCACTCCCCAAGGAAGAGCCAATGTTTTCCACCCATAATCCTTTCTGCCGACCCCTAGTTCCCTCT
GCTCAGCCAAGCTTGTTATCAGCTTTCAGGGCCATGGTTCACATTAGAATAAAAGGTAGTAATAGAAAAA
AAAAAAAAACCCGGGCCTTATGGCGGGGGGGAAAGGGGACAGGAAGGGCATGGCTAGGGTTCGGCCGGTC
TTGACAGGGGATTGCGGGCCA

>DEX0356_032.nt.1
GTGGATCCCCCGGGCTGCAGGCGGCACGAGGCCGCTGCGAAGGGAGCCGCCGCCATGTCTGCGCATCTGC
AATGGATGGTCGTGCGGAACTGCTCCAGTTTCCTGATCAAGAGGAATAAGCAGACCTACAGCACTGAGCC
CAATAACTTGAAGGCCCGCAATTCCTTCCGCTACAACGGACTGATTCACCGCAAGACTGTGGGCGTGGAG
CCGGCAGCCGACGGCAAAGGTGTCGTGGTGGTCATTAAGCGGAGATCCGGCCAGCGGAAGCCTGCCACCT
CCTATGTGCGGACCACCATCAACAAGAATGCTCGCGCCACGCTCAGCAGCATCAGACACATGATCCGCAA
GAACAAGTACCGCCCCGACCTGCGCATGGCAGCCATCCGCAGGGCCAGCGCCATCCTGCGCAGCCAGAAG
CCTGTGATGGTGAAGAGGAAGCGGACCCGCCCCACCAAGAGCTCCTGAGCCCCCTGCCCCCAGAGCAATA
AAGTCAGCTGGCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAR
AAAAAAAAAAAAAAAAAAAAAAAAGGGGGGGGGGGGGGCAAAAAAAACCCCCGGGGGGGGCCCAACATATAG
CGACCACACACCTTCGTGGGAGAGAAGAGGCCGCCACACAGCGAGGGTCATACAACAAAACAGCGGGCGG
CGCGCCGCAAAAAAACACGGCGCGGGGGGCACAAACGTCAGCGTGGCTATCTGGGAGAAGACCCCTGCCG
GCGGGGAGCGACACAGCGAGACCCCCCCGAGCAAGAGACGCGGGGCGAAAAACTGGTCCGTTGGCGCACA
CCCGCAGTCGGCGCGAACGCGCCGGTTACAAA

>DEX0356_033.nt.1
CACCAGAAGTGAGGACGGCTGCCTGGGCCCTGCCGTGTGCTGCTGGGGGGGCTGGCTCCATGGAGGGGGT
CCTCTGCCCTGCTGGCTGGTGGCTGAGGCCCCAGGCCTGTGCCCAGGGCCGCTGGCAGCACAGTGAGGCC
ACGTCCACAGCAGAGTTGGCCGTGGAGATAGTGTACGTGGTGCCCTCCGTAGCTGAGAGCCTCCCAGCCC
CACACTGGCTGAGTACCCAGGGCTTCCAGAATTTGGAAAAAGGAGCAGTGTTTTGGTCGTGGAGCTGTGG
GGTGCCGTGCGCGTCCATGGGCTGGGTCTGCGCTCAGGCTGCCCTGCCGTGGAAGTGTGGGTGTGGGTAT
GGGTTGGGGGTCTGGGACCCCAGGCAGCCCAAGTGGGAGCAGGGCCGGCCGGTGGGGAAGGGAGGCTCTG
GCCTGGTGGGAAGTGCAGCTCCGCGTTGTCCCTTCTCTTGCAGCGGGGATCTGACGAGCTTTTCTCTACT
TGTGTCACTAACGGACCGTTTATCATGAGCAGCAACTCGGCTTCTGCAGCAAACGGAAATGACAGCAAGA

AGTTCAAAGGTGACAGCCGAAGTGCAGGCGTCCCCTCTAGAGTGATCCACATCCGGAAGCTCCCCATCGA
CGTCACGGAGGGGGAAGTCATCTCCCTGGGGCTGCCCTTTGGGAAGGTCACCAACCTCCTGATGCTGAAG
GGGAAAAACCAGGCCTTCATCGAGATGAACACGGAGGAGGCTGCCAACACCATGGTGAACTACTACACCT
CGGTGACCCCTGTGCTGCGCGGCCAGCCCATCTACATCCAGTTTCTCTGTGCTCTTTCTACCGCCCCCGC
GTCCTGTCCCGGGGGCTCTCCTAGGATCCCCTTTCCGTAAAAGCGTGTAACAAGGGTGTAAATATTTATA
ATTTTTTATACCTGTTGTGAGACCCGAGGGGCGGCGGCGCGGTTTTTTATGGTGACACAAATGTATATTT
TGCTAACAGCAATTCCAGGCTCAGTATTGTGACCGCGGAGCCACAGGGGACCCCACGCACATTCCGTTGC
CTTACCCGATGGCTTGTGACGCGGAGAGAACCGATTAAAACCGTTTGAGAAACTCCTCCCTTGTCTAGCC
CTGTGTTCGCTGTGGACGCTGTAGAGGCAGGTTGGCCAGTCTGTACCTGGACTTCGAATAAATCTTCTGT
ATCCTCGCTCCGTTCCGCCTT

>DEX0356_033.nt.2
CACCGTGGTGTTGGCAGCCTCCTCCGTGTTCATCTCGATGAAGGCCTGGTTTTTCCCCTTCAGCATCAGG
AGGTTGGTGACCTTCCCAAAGGGCAGCCCCAGGGAGATGACTTCCCCCTCCGTGACGTCGATGGGGCTAT
TCCGGCGCCTCCACTCCGTCCCCCGCGGGTCTGCTCTGTGTGCCATGGACGGCATTGTCCCAGATATAGC
CGTTGGTACAAAGCGGGGATCTGACGAGCTTTTCTCTACTTGTGTCACTAACGGACCGTTTATCATGAGC
AGCAACTCGGCTTCTGCAGCAAACGGAAATGACAGCAAGAAGTTCAAAGGTGACAGCCGAAGTGCAGGCG
TCCCCTCTAGAGTGATCCACATCCGGAAGCTCCCCATCGACGTCACGGAGGGGGAAGTCATCTCCCTGGG
GCTGCCCTTTGGGAAGGTCACCAACCTCCTGATGCTGAAGGGGAAAAACCAGGTACCTGAGCCGCGTTTC
TCCGGGGTGCTCACACCGTGCAGGCGGGGACGAGGAGGGCCCAGCGCTCACTGCCTCCCCAACAGGCCTT
CATCGAGATGAACACGGAGGAGGCTGCCAACACCATGGTGAACTACTACACCTCGGTGACCCCTGTGCTG
CGCGGCCAGCCCATCTACATCCAGTTCTCCAACCACAAGGAGCTGAAGACCGACAGCTCTCCCAACCAGG
CGCGGGCCCAGGCGGCCCTGCAGGCGGTGAACTCGGTCCAGTCGGGGAACCTGGCCTTGGCTGCCTCGGC
GGCGGCCGTGGACGCAGGGATGGCGATGGCCGGGCAGAGCCCCGTGCTCAGGATCATCGTGGAGAACCTC
TTCTACCCTGTGACCCTGGATGTGCTGCACCAGATTTTCTCCAAGTTCGGCACAGTGTTGAAGATCATCA
CCTTCACCAAGAACAACCAGTTCCAGGCCCTGCTGCAGTATGCGGACCCCGTGAGCGCCCAGCACGCCAA
GCTGTCGCTGGACGGGCAGAACATCTACAACGCCTGCTGCACGCTGCGCATCGACTTTTCCAAGCTCACC
AGCCTCAACGTCAAGTACAACAATGACAAGAGCCGTGACTACACACGCCCAGACCTGCCTTCCGGGGACA
GCCAGCCCTCGCTGGACCAGACCATGGCCGCGGCCTTCGCCTCTCCGTATGCAGGAGCTGGTTTCCCTCC
CACCTTTGCCATTCCTCAAGCTGCAGGGTATTCAAACGCTTGGTCTTGGTTCCCCAGCGACTGCATGCCC
ACACCACCTTCCCAGGCAGCTCCGCATCCACGGCGGCAGCCTGGGCGGACTGGGCACTCGAGTGCCAGGT
CAGGGGCCCTTCCCGGGAGAGGGGACGCCACGTCCACAGAGCAGGCTTGGCCAGGGCAGTGGTAGGACAG
GGCTGTGGAGGCCCACGTGTGGACGGCGCCAGCCAGAAGCCGCTAATCGCACAGTCTTTGTGGTCTGGTT
CCCTTCAAGCACCTCCGGGCGAGCGCGAGGACCGCCAGTGGTTGGAGGGATGTCTCTAGTAGTTGAATTT
GAGTGGCCTGGTAAGCGCGCGGCCCGAGGCTGAGCAGCATGACCGGCGGCGGGACGTGTGTGCGTGGCCG
TCCTCCCGCTTCCCTTCTGGGACCGGGGCCCGGCCGCCCCAGTGCTGGCCGTGGACGGAGCAGCGCGTGC
CGCCCGGCGGCCGCCTCGTCTGCATGGAGGCATGGGCGCGCATGCGCGGTGGTCCCGGGACGGGCCCTGC
TTGCCGAGGGCCCCGTGTCTGTGCTGGCGGAGCCGGAGCTTTTCTGTGCTGTGGGCGGGGCGCATGCAGA
TGAGCCCAGGCCCGGCCCGGCCCGTGCTGTGAGGAGAGGCGGGCGCTGGTGCAGGAGCCGGCGGGTGGCT
GTGCGGGGGTCGGGCGACCCCACAGGCACCCAGGGTAGGGCCAGAGCCAGGGCCGCCTCCCGCGCGGCTC
GGCTCCTCGGTGCATGAGGACGGGGAGCGTCGGCCTCTCCCACTCTGCGGTGGAGTCGGGGGCGCCGCCG
CTCATCTCACCTCCTGCTTTTCCAGGCCTTTCCGTTCCGAACGTCCACGGCGCCCTGGCCCCCCTGGCCA
TCCCCTCGGCGGCGGCGGCAGCTGCGGCGGCAGGTCGGATCGCCATCCCGGGCCTGGCGGGGGCAGGAAA
TTCTGTATTGCTGGTCAGCAACCTCAACCCAGAGGTACGTGGGCTTTTCCTCCGCGCCGCCGTTCCTCCC
GGAAGAGCGAGCAGGGGATGTTGTGCTCGGGCTCGGGTCCCGGAGCAGCGCCGCCCCTCGCTGTGTCTGC
GCCTCTGCCCCGGCAGCCCCTCTGCTGCAGCGCTGAGACCCTCCTTTTCCAAGATGGCTTGAGTTTTCCT
CTTTTCCTGAATTCACATCTTGGTTCGCGTTTTCTCTTGCATGATACGCAGAATGAATTATTTTTTGGTT
ACCCAAAGGCAATCTAAAAATTACTTGGAGCTGAGCAGGGCGCGCAGTTCAGAGCCCCAGCAGGCCGCCC
TGCTGGTCAGCGGGACACACAGGGTTAGCGCGGCGGGGTGTGGGCAGCCCCTGTGGCTCGAGAGCCTGTC
ACTTAACTGGTGGATCCGTTATTTCCCAGAATGGTTAAATTCCCCTCGGGGAGAGCTTGGGTGAAACAAG
AGGCCCAAAATATGTAACTCTTGATGGATGGGATCCTCCTGGAATCAGCATTTCAGGCTGTTCTGTTTCC
CGGCAACTCCTGTGACTTCCTCAAGGCGGTGCCCAGCCTGGCCCTCCCTCGCGGGGGCCCAGGACGCACT
CCCTGCCTGTGCCTCTTACCACGAGCCGGGTGCTGTCCCTCCAGCTCCTCCTGAGCGAGGCCGTGCACTA
ACAGACTCGATCTCCCTTCACGCCTGTAATCGTCCTTAGAGACTGACTTGTGCCACAGCTGCGGGAGACA
GCGCTGCTGCAGGGCTTGGCTGAGGTGGCCGCGAGCGGAGCTGCACGCGCGCTCCTGGGGGGCGTGGTTG
GGATCGGCTTCTGGGAGCCCCTCCAGCAGCGCTGTCTCTGCCGGCATGTGAGTGAAGGTGCTTCTCTTAC
CGTGTGCCCGTCGAGGTAAACCTGTAACTGGAATGTGTGTGGAGTGTGACTGATAGAACACTACCTGATT
CTTATGTATTTACTGACCTGTGTTTTTTTGCTACTTTTTTTCTTTTCTCCCCTTCCCCTTTCCCTATTTT
TTTTCTTGCCCTGATCCGGAATTTCTTTGCCAACTGACTGCACGGTACTTCTGCTTCCTGTTGTTGCTTG
AAACAAAACAAAAACATAAACAAATAAAAAACAAAAATTCCCCCTCAAACCCTGCTCTCCGGAAACCAAC

CTGCCCTTGAATATTAACATCCTGACAACTTCATCATCCATCAACCACTGCACGCCTGCGGGGACTGTCT
TCCTCGTGTGGACGATTGGCAACTCGCCCCCCTTGACCTCTCCCTCTCCCCTGTCCCTCCGCTGCCTTGC
TCTGCTGTCTCTAAAGAGAGTCACACCCCAAAGCCTCTTTATTCTTTTCGGCGTCTACGGTGACGTGCAG
CGCGTGAAGATCCTGTTCAATAAGAAGGAGAACGCCCTAGTGCAGATGGCGGACGGCAACCAGGCCCAGC
TGGCCATGAGCCACCTGAACGGGCACAAGCTGCACGGGAAGCCCATCCGCATCACGCTCTCGAAGCACCA
GAACGTGCAGCTGCCCCGCGAGGGCCAGGAGGACCAGGGCCTGACCAAGGACTACGGCAACTCACCCCTG
CACCGCTTCAAGAAGCCGGGCTCCAAGAACTTCCAGAACATATTCCCGCCCTCGGCCACGCTGCACCTCT
CCAACATCCCGCCCTCAGTCTCCGAGGAGGATCTCAAGGTCCTGTTTTCCAGCAATGGGGGCGTCGTCAA
AGGATTCAAGTTCTTCCAGAAGGACCGCAAGATGGCACTGATCCAGATGGGCTCCGTGGAGGAGGCGGTC
CAGGCCCTCATTGACCTGCACAACCACGACCTCGGGGAGAACCACCACCTGCGGGTCTCCTTCTCCAAGT
CCACCATCTAGGGGCACAGGCCCCCACGGCCGGGCCCCCTGGCGACAACTTCCATCATTCCAGAGAAAAG
CCACTTTAAAAACAGCTGAAGTGACCTTAGCAGACCAGAGATTTTATTTTTTTAAAGAGAAATCAGTTTA
CCTGTTTTTAAAAAAATTAAATCTAGTTCACCTTGCTCACCCTGCGGTGACAGGGACAGCTCAGGCTCTT
GGTGACTGTGGCAGCGGGAGTTCCCGGCCCTCCACACCCGGGGCCAGACCCTCGGGGCCATGCCTTGGTG
GGGCCTGTGTCGGGCGTGGGGCCTGCAGGTGGGCGCCCCGACCACGACTTGGCTTCCTTGTGCCTTAAAA
AACCTGCCTTCCTGCAGCCACACACCCACCCGGGGTGTCCTGGGGACCCAAGGGGTGGGGGGGTCACACC
AGAGAGAGGCAGGGGGCCTGGCCGGCTCCTGCAGGATCATGCAGCTGGGGCGCGGCGGCCGCGGCTGCGA
CACCCCAACCCCAGCCCTCTAATCAAGTCACGTGATTCTCCCTTCACCCCGCCCCCAGGGCCTTCCCTTC
TGCCCCCAGGCGGGCTCCCCGCTGCTCCAGCTGCGGAGCTGGTCGACATAATCTCTGTATTATATACTTT
GCAGTTGCAGACGTCTGTGCCTAGCAATATTTCCAGTTGACCAAATATTCTAATCTTTTTTCATTTATAT
GCAAAAGAAATAGTTTTAAGTAACTTTTTATAGCAAGATGATACAATGGTATGAGTGTAATCTAAACTTC
CTTGTGGTATTACCTTGTATGCTGTTACTTTTATTTTATTCCTTGTAATTAAGTCACAGGCAGGACCCAG
TTTCCAGAGAGCAGGCGGGGCCGCCCAGTGGGTCAGGCACAGGGAGCCCCGGTCCTATCTTAGAGCCCCT
GAGCTTCAGGGAAGGGGCGGGCGTGTCGCCGCCTCTGGCATCGCCTCCGGTTGCCTTACACCACGCCTTC
ACCTGCAGTCGCCTAGAAAACTTGCTCTCAAACTTCAGGGTTTTTTCTTCCTTCAAATTTTGGACCAAAG
TCTCATTTCTGTGTTTTGCCTGCCTCTGATGCTGGGACCCGGAAGGCGGGCGCTCCTCCTGTCTTCTCTG
TGCTCTTTCTACCGCCCCCGCGTCCTGTCCCGGGGGCTCTCCTAGGATCCCCTTTCCGTAAAAGCGTGTA
ACAAGGGTGTAAATATTTATAATTTTTTATACCTGTTGTGAGACCCGAGGGGCGGCGGCGCGGTTTTTTA
TGGTGACACAAATGTATATTTTGCTAACAGCAATTCCAGGCTCAGTATTGTGACCGCGGAGCCACAGGGG
ACCCCACGCACATTCCGTTGCCTTACCCGATGGCTTGTGACGCGGAGAGAACCGATTAAAACCGTTTGAG
AAACTCCTCCCTTGTCTAGCCCTGTGTTCGCTGTGGACGCTGTAGAGGCAGGTTGGCCAGTCTGTACCTG
GACTTCGAATAAATCTTCTGTATCCTCGCTCCGTTCCGCCTT

>DEX0356_033.nt.3

CACCGTGGTGTTGGCAGCCTCCTCCGTGTTCATCTCGATGAAGGCCTGGTTTTTCCCCTTCAGCATCAGG
AGGTTGGTGACCTTCCCAAAGGGCAGCCCCAGGGAGATGACTTCCCCCTCCGTGACGTCGATGGGGCTAT
TCCGGCGCCTCCACTCCGTCCCCCGCGGGTCTGCTCTGTGTGCCATGGACGGCATTGTCCCAGATATAGC
CGTTGGTACAAAGCGGGGATCTGACGAGCTTTTCTCTACTTGTGTCACTAACGGACCGTTTATCATGAGC
AGCAACTCGGCTTCTGCAGCAAACGGAAATGACAGCAAGAAGTTCAAAGGTGACAGCCGAAGTGCAGGCG
TCCCCTCTAGAGTGATCCACATCCGGAAGCTCCCCATCGACGTCACGGAGGGGGAAGTCATCTCCCTGGG
GCTGCCCTTTGGGAAGGTCACCAACCTCCTGATGCTGAAGGGGAAAAACCAGGTACCTGAGCCGCGTTTC
TCCGGGGTGCTCACACCGTGCAGGCGGGGACGAGGAGGGCCCAGCGCTCACTGCCTCCCCAACAGGCCTT
CATCGAGATGAACACGGAGGAGGCTGCCAACACCATGGTGAACTACTACACCTCGGTGACCCCTGTGCTG
CGCGGCCAGCCCATCTACATCCAGTTCTCCAACCACAAGGAGCTGAAGACCGACAGCTCTCCCAACCAGG
CGCGGGCCCAGGCGGCCCTGCAGGCGGTGAACTCGGTCCAGTCGGGGAACCTGGCCTTGGCTGCCTCGGC
GGCGGCCGTGGACGCAGGGATGGCGATGGCCGGGCAGAGCCCCGTGCTCAGGATCATCGTGGAGAACCTC
TTCTACCCTGTGACCCTGGATGTGCTGCACCAGATTTTCTCCAAGTTCGGCACAGTGTTGAAGATCATCA
CCTTCACCAAGAACAACCAGTTCCAGGCCCTGCTGCAGTATGCGGACCCCGTGAGCGCCCAGCACGCCAA
GCTGTCGCTGGACGGGCAGAACATCTACAACGCCTGCTGCACGCTGCGCATCGACTTTTCCAAGCTCACC
AGCCTCAACGTCAAGTACAACAATGACAAGAGCCGTGACTACACACGCCCAGACCTGCCTTCCGGGGACA
GCCAGCCCTCGCTGGACCAGACCATGGCCGCGGCCTTCGGGTAAGAGGCTGCCCGACGCGGCGCCAGTGT
GCAGAGTGGTCTATTAGGGCCGCTCAGTCCGGAGCCCCGGCGGCACGGGCCCGGCCCTGCACCAGGGTGA
TGCACCTGCTGCTCTCTGCACGGCCAGCACAGCACGGTCCAGTGTCCCCCACGTCGGGACCACGGCCCCC
CTGGAGCAGCGGATCTGCCCGCGAAGGCTCTGCCGGGGCCGCCCGCCGGCCGGGTTGGGGCCCATCCCGC
AGCACAGCGCCCGCTCGCGGTGGAGGTTGTGGGTGCGATGATTAGTGTCTCATTTATTTCTAGGTGCACC
TGGTATAATCTCAGCCTCTCCGTATGCAGGAGCTGGTTTCCCTCCCACCTTTGCCATTCCTCAAGCTGCA
GGGTATTCAAACGCTTGGTCTTGGTTCCCCAGCGACTGCATGCCCACACCACCTTCCCAGGCAGCTCCGC
ATCCACGGCGGCAGCCTGGGCGGACTGGGCACTCGAGTGCCAGGTCAGGGGCCCTTCCCGGGAGAGGGGA
CGCCACGTCCACAGAGCAGGCTTGGCCAGGGCAGTGGTAGGACAGGGCTGTGGAGGCCCACGTGTGGACG
GCGCCAGCCAGAAGCCGCTAATCGCACAGTCTTTGTGGTCTGGTTCCCTTCAAGCACCTCCGGGCGAGCG

CGAGGACCGCCAGTGGTTGGAGGGATGTCTCTAGTAGTTGAATTTGAGTGGCCTGGTAAGCGCGCGGCCC
GAGGCTGAGCAGCATGACCGGCGGCGGGACGTGTGTGCGTGGCCGTCCTCCCGCTTCCCTTCTGGGACCG
GGGCCCGGCCGCCCCAGTGCTGGCCGTGGACGGAGCAGCGCGTGCCGCCCGGCGGCCGCCTCGTCTGCAT
GGAGGCATGGGCGCGCATGCGCGGTGGTCCCGGGACGGGCCCTGCTTGCCGAGGGCCCCGTGTCTGTGCT
GGCGGAGCCGGAGCTTTTCTGTGCTGTGGGCGGGGCGCATGCAGATGAGCCCAGGCCCGGCCCGGCCCGT
GCTGTGAGGAGAGGCGGGCGCTGGTGCAGGAGCCGGCGGGTGGCTGTGCGGGGGTCGGGCGACCCCACAG
GCACCCAGGGTAGGGCCAGAGCCAGGGCCGCCTCCCGCGCGGCTCGGCTCCTCGGTGCATGAGGACGGGG
AGCGTCGGCCTCTCCCACTCTGCGGTGGAGTCGGGGGCGCCGCCGCTCATCTCACCTCCTGCTTTTCCAG
GCCTTTCCGTTCCGAACGTCCACGGCGCCCTGGCCCCCCTGGCCATCCCCTCGGCGGCGGCGGCAGCTGC
GGCGGCAGGTCGGATCGCCATCCCGGGCCTGGCGGGGGCAGGAAATTCTGTATTGCTGGTCAGCAACCTC
AACCCAGAGGTACGTGGGCTTTTCCTCCGCGCCGCCGTTCCTCCCGGAAGAGCGAGCAGGGGATGTTGTG
CTCGGGCTCGGGTCCCGGAGCAGCGCCGCCCCTCGCTGTGTCTGCGCCTCTGCCCCGGCAGCCCCTCTGC
TGCAGCGCTGAGACCCTCCTTTTCCAAGATGGCTTGAGTTTTCCTCTTTTCCTGAATTCACATCTTGGTT
CGCGTTTTCTCTTGCATGATACGCAGAATGAATTATTTTTTGGTTACCCAAAGGCAATCTAAAAATTACT
TGGAGCTGAGCAGGGCGCGCAGTTCAGAGCCCCAGCAGGCCGCCCTGCTGGTCAGCGGGACACACAGGGT
TAGCGCGGCGGGGTGTGGGCAGCCCCTGTGGCTCGAGAGCCTGTCACTTAACTGGTGGATCCGTTATTTC
CCAGAATGGTTAAATTCCCCTCGGGGAGAGCTTGGGTGAAACAAGAGGCCCAAAATATGTAACTCTTGAT
GGATGGGATCCTCCTGGAATCAGCATTTCAGGCTGTTCTGTTTCCCGGCAACTCCTGTGACTTCCTCAAG
GCGGTGCCCAGCCTGGCCCTCCCTCGCGGGGGCCCAGGACGCACTCCCTGCCTGTGCCTCTTACCACGAG
CCGGGTGCTGTCCCTCCAGCTCCTCCTGAGCGAGGCCGTGCACTAACAGACTCGATCTCCCTTCACGCCT
GTAATCGTCCTTAGAGACTGACTTGTGCCACAGCTGCGGGAGACAGCGCTGCTGCAGGGCTTGGCTGAGG
TGGCCGCGAGCGGAGCTGCACGCGCGCTCCTGGGGGGGCGTGGTTGGGATCGGCTTCTGGGAGCCCCTCCA
GCAGCGCTGTCTCTGCCGGCATGTGAGTGAAGGTGCTTCTCTTACCGTGTGCCCGTCGAGGTAAACCTGT
AACTGGAATGTGTGTGGAGTGTGACTGATAGAACACTACCTGATTCTTATGTATTTACTGACCTGTGTTT
TTTTGCTACTTTTTTTCTTTTCTCCCCTTCCCCTTTCCCTATTTTTTTTCTTGCCCTGATCCGGAATTTC
TTTGCCAACTGACTGCACGGTACTTCTGCTTCCTGTTGTTGCTTGAAACAAAACAAAAACATAAACAAAT
AAAAAACAAAAATTCCCCCTCAAACCCTGCTCTCCGGAAACCAACCTGCCCTTGAATATTAACATCCTGA
CAACTTCATCATCCATCAACCACTGCACGCCTGCGGGGACTGTCTTCCTCGTGTGGACGATTGGCAACTC
GCCCCCCTTGACCTCTCCCTCTCCCCTGTCCCTCCGCTGCCTTGCTCTGCTGTCTCTAAAGAGAGTCACA
CCCCAAAGCCTCTTTATTCTTTTCGGTATGTTATCGTTCACACTTTTATTACCTTGTTTTCATTAATTGA
GATGATTTTGATGCCCTGAGACGTATTATGAGTTTGCGGTTTGGCACTCTGATGCTCCGTGGCATCCGCC
TCGTTTTATGGTTTGCTTTCGGTTTGCGAATTTTATTTGGTCCCGTAGATACGTACGCATGGTTTATCGC
CCTGCATGCTTTTCAGAGTTAGACCTGTCGGTGGCATATGCCACCGTGGCCACCCGCTGGCAGCTTACCT
GTCCTGGATGCTATGACTTTGCTGAACGGAGCTGCTCCTGTTAGCGCGCCCTGTGGCTGCGAGACGCAGC
TCCGCAGTGGCCGATAAAGCAAACCCGGCCGGGCTGAGCCGGGCCTTGTGGGGGTGCGCGGGGCCGGGGC
TGACGGGGAGATGGGCGGGGCAGGCAGCAGGAGACTCAGGCCCCATCCCTGGGCTTTTGAAGGCGCGTCT
ACGGTGACGTGCAGCGCGTGAAGATCCTGTTCAATAAGAAGGAGAACGCCCTAGTGCAGATGGCGGACGG
CAACCAGGCCCAGCTGGCCATGAGCCACCTGAACGGGCACAAGCTGCACGGGAAGCCCATCCGCATCACG
CTCTCGAAGCACCAGAACGTGCAGCTGCCCCGCGAGGGCCAGGAGGACCAGGGCCTGACCAAGGACTACG
GCAACTCACCCCTGCACCGCTTCAAGAAGCCGGGCTCCAAGAACTTCCAGAACATATTCCCGCCCTCGGC
CACGCTGCACCTCTCCAACATCCCGCCCTCAGTCTCCGAGGAGGATCTCAAGGTCCTGTTTTCCAGCAAT
GGGGGCGTCGTCAAAGGATTCAAGTTCTTCCAGAAGGACCGCAAGATGGCACTGATCCAGATGGGCTCCG
TGGAGGAGGCGGTCCAGGCCCTCATTGACCTGCACAACCACGACCTCGGGGAGAACCACCACCTGCGGGT
CTCCTTCTCCAAGTCCACCATCTAGGGGCACAGGCCCCCACGGCCGGGCCCCCTGGCGACAACTTCCATC
ATTCCAGAGAAAAGCCACTTTAAAAACAGCTGAAGTGACCTTAGCAGACCAGAGATTTTATTTTTTTAAA
GAGAAATCAGTTTACCTGTTTTTAAAAAAATTAAATCTAGTTCACCTTGCTCACCCTGCGGTGACAGGGA
CAGCTCAGGCTCTTGGTGACTGTGGCAGCGGGAGTTCCCGGCCCTCCACACCCGGGGCCAGACCCTCGGG
GCCATGCCTTGGTGGGGCCTGTGTCGGGCGTGGGGCCTGCAGGTGGGCGCCCCGACCACGACTTGGCTTC
CTTGTGCCTTAAAAAACCTGCCTTCCTGCAGCCACACACCCACCCGGGGTGTCCTGGGGACCCAAGGGGT
GGGGGGGTCACACCAGAGAGAGGCAGGGGGCCTGGCCGGCTCCTGCAGGATCATGCAGCTGGGGCGCGGC
GGCCGCGGCTGCGACACCCCAACCCCAGCCCTCTAATCAAGTCACGTGATTCTCCCTTCACCCCGCCCCC
AGGGCCTTCCCTTCTGCCCCCAGGCGGGCTCCCCGCTGCTCCAGCTGCGGAGCTGGTCGACATAATCTCT
GTATTATATACTTTGCAGTTGCAGACGTCTGTGCCTAGCAATATTTCCAGTTGACCAAATATTCTAATCT
TTTTTCATTTATATGCAAAAGAAATAGTTTTAAGTAACTTTTTATAGCAAGATGATACAATGGTATGAGT
GTAATCTAAACTTCCTTGTGGTATTACCTTGTATGCTGTTACTTTTATTTTATTCCTTGTAATTAAGTCA
CAGGCAGGACCCAGTTTCCAGAGAGCAGGCGGGGCCGCCCAGTGGGTCAGGCACAGGGAGCCCCGGTCCT
ATCTTAGAGCCCCTGAGCTTCAGGGAAGGGGCGGGCGTGTCGCCGCCTCTGGCATCGCCTCCGGTTGCCT
TACACCACGCCTTCACCTGCAGTCGCCTAGAAAACTTGCTCTCAAACTTCAGGGTTTTTTCTTCCTTCAA
ATTTTGGACCAAAGTCTCATTTCTGTGTTTTGCCTGCCTCTGATGCTGGGACCCGGAAGGCGGGCGCTCC
TCCTGTCTTCTCTGTGCTCTTTCTACCGCCCCCGCGTCCTGTCCCGGGGGCTCTCCTAGGATCCCCTTTC

CGTAAAAGCGTGTAACAAGGGTGTAAATATTTATAATTTTTTATACCTGTTGTGAGACCCGAGGGGCGGC
GGCGCGGTTTTTTATGGTGACACAAATGTATATTTTGCTAACAGCAATTCCAGGCTCAGTATTGTGACCG
CGGAGCCACAGGGGACCCCACGCACATTCCGTTGCCTTACCCGATGGCTTGTGACGCGGAGAGAACCGAT
TAAAACCGTTTGAGAAACTCCTCCCTTGTCTAGCCCTGTGTTCGCTGTGGACGCTGTAGAGGCAGGTTGG
CCAGTCTGTACCTGGACTTCGAATAAATCTTCTGTATCCTCGCTCCGTTCCGCCTT

>DEX0356_033.nt.4
CACCGTGGTGTTGGCAGCCTCCTCCGTGTTCATCTCGATGAAGGCCTGGTTTTTCCCCTTCAGCATCAGG
AGGTTGGTGACCTTCCCAAAGGGCAGCCCCAGGGAGATGACTTCCCCCTCCGTGACGTCGATGGGGCTAT
TCCGGCGCCTCCACTCCGTCCCCCGCGGGTCTGCTCTGTGTGCCATGGACGGCATTGTCCCAGATATAGC
CGTTGGTACAAAGCGGGGATCTGACGAGCTTTTCTCTACTTGTGTCACTAACGGACCGTTTATCATGAGC
AGCAACTCGGCTTCTGCAGCAAACGGAAATGACAGCAAGAAGTTCAAAGGTGACAGCCGAAGTGCAGGCG
TCCCCTCTAGAGTGATCCACATCCGGAAGCTCCCCATCGACGTCACGGAGGGGGAAGTCATCTCCCTGGG
GCTGCCCTTTGGGAAGGTCACCAACCTCCTGATGCTGAAGGGGAAAAACCAGGTACCTGAGCCGCGTTTC
TCCGGGGTGCTCACACCGTGCAGGCGGGGACGAGGAGGGCCCAGCGCTCACTGCCTCCCCAACAGGCCTT
CATCGAGATGAACACGGAGGAGGCTGCCAACACCATGGTGAACTACTACACCTCGGTGACCCCTGTGCTG
CGCGGCCAGCCCATCTACATCCAGTTCTCCAACCACAAGGAGCTGAAGACCGACAGCTCTCCCAACCAGG
CGCGGGCCCAGGCGGCCCTGCAGGCGGTGAACTCGGTCCAGTCGGGGAACCTGGCCTTGGCTGCCTCGGC
GGCGGCCGTGGACGCAGGGATGGCGATGGCCGGGCAGAGCCCCGTGCTCAGGATCATCGTGGAGAACCTC
TTCTACCCTGTGACCCTGGATGTGCTGCACCAGATTTTCTCCAAGTTCGGCACAGTGTTGAAGATCATCA
CCTTCACCAAGAACAACCAGTTCCAGGCCCTGCTGCAGTATGCGGACCCCGTGAGCGCCCAGCACGCCAA
GCTGTCGCTGGACGGGCAGAACATCTACAACGCCTGCTGCACGCTGCGCATCGACTTTTCCAAGCTCACC
AGCCTCAACGTCAAGTACAACAATGACAAGAGCCGTGACTACACACGCCCAGACCTGCCTTCCGGGGACA
GCCAGCCCTCGCTGGACCAGACCATGGCCGCGGCCTTCGGCCTTTCCGTTCCGAACGTCCACGGCGCCCT
GGCCCCCCTGGCCATCCCCTCGGCGGCGGCGGCAGCTGCGGCGGCAGGTCGGATCGCCATCCCGGGCCTG
GCGGGGGCAGGAAATTCTGTATTGCTGGTCAGCAACCTCAACCCAGAGGCGTCTACGGTGACGTGCAGCG
CGTGAAGATCCTGTTCAATAAGAAGGAGAACGCCCTAGTGCAGATGGCGGACGGCAACCAGGCCCAGCTG
GCCATGAGCCACCTGAACGGGCACAAGCTGCACGGGAAGCCCATCCGCATCACGCTCTCGAAGCACCAGA
ACGTGCAGCTGCCCCGCGAGGGCCAGGAGGACCAGGGCCTGACCAAGGACTACGGCAACTCACCCCTGCA
CCGCTTCAAGAAGCCGGGCTCCAAGAACTTCCAGAACATATTCCCGCCCTCGGCCACGCTGCACCTCTCC
AACATCCCGCCCTCAGTCTCCGAGGAGGATCTCAAGGTCCTGTTTTCCAGCAATGGGGGCGTCGTCAAAG
GATTCAAGTTCTTCCAGAAGGACCGCAAGATGGCACTGATCCAGATGGGCTCCGTGGAGGAGGCGGTCCA
GGCCCTCATTGACCTGCACAACCACGACCTCGGGGAGAACCACCACCTGCGGGTCTCCTTCTCCAAGTCC
ACCATCTAGGGGCACAGGCCCCCACGGCCGGGCCCCCTGGCGACAACTTCCATCATTCCAGAGAAAAGCC
ACTTTAAAAACAGCTGAAGTGACCTTAGCAGACCAGAGATTTTATTTTTTTAAAGAGAAATCAGTTTACC
TGTTTTTAAAAAAATTAAATCTAGTTCACCTTGCTCACCCTGCGGTGACAGGGACAGCTCAGGCTCTTGG
TGACTGTGGCAGCGGGAGTTCCCGGCCCTCCACACCCGGGGCCAGACCCTCGGGGCCATGCCTTGGTGGG
GCCTGTGTCGGGCGTGGGGCCTGCAGGTGGGCGCCCCGACCACGACTTGGCTTCCTTGTGCCTTAAAAAA
CCTGCCTTCCTGCAGCCACACACCCACCCGGGGTGTCCTGGGGACCCAAGGGGTGGGGGGGTCACACCAG
AGAGAGGCAGGGGGCCTGGCCGGCTCCTGCAGGATCATGCAGCTGGGGCGCGGCGGCCGCGGCTGCGACA
CCCCAACCCCAGCCCTCTAATCAAGTCACGTGATTCTCCCTTCACCCCGCCCCCAGGGCCTTCCCTTCTG
CCCCCAGGCGGGCTCCCCGCTGCTCCAGCTGCGGAGCTGGTCGACATAATCTCTGTATTATATACTTTGC
AGTTGCAGACGTCTGTGCCTAGCAATATTTCCAGTTGACCAAATATTCTAATCTTTTTTCATTTATATGC
AAAAGAAATAGTTTTAAGTAACTTTTTATAGCAAGATGATACAATGGTATGAGTGTAATCTAAACTTCCT
TGTGGTATTACCTTGTATGCTGTTACTTTTATTTTATTCCTTGTAATTAAGTCACAGGCAGGACCCAGTT
TCCAGAGAGCAGGCGGGGCCGCCCAGTGGGTCAGGCACAGGGAGCCCCGGTCCTATCTTAGAGCCCCTGA
GCTTCAGGGAAGGGGCGGGCGTGTCGCCGCCTCTGGCATCGCCTCCGGTTGCCTTACACCACGCCTTCAC
CTGCAGTCGCCTAGAAAACTTGCTCTCAAACTTCAGGGTTTTTTCTTCCTTCAAATTTTGGACCAAAGTC
TCATTTCTGTGTTTTGCCTGCCTCTGATGCTGGGACCCGGAAGGCGGGCGCTCCTCCTGTCTTCTCTGTG
CTCTTTCTACCGCCCCCGCGTCCTGTCCCGGGGGCTCTCCTAGGATCCCCTTTCCGTAAAAGCGTGTAAC
AAGGGTGTAAATATTTATAATTTTTTATACCTGTTGTGAGACCCGAGGGGCGGCGGCGCGGTTTTTTATG
GTGACACAAATGTATATTTTGCTAACAGCAATTCCAGGCTCAGTATTGTGACCGCGGAGCCACAGGGGAC
CCCACGCACATTCCGTTGCCTTACCCGATGGCTTGTGACGCGGAGAGAACCGATTAAAACCGTTTGAGAA
ACTCCTCCCTTGTCTAGCCCTGTGTTCGCTGTGGACGCTGTAGAGGCAGGTTGGCCAGTCTGTACCTGGA
CTTCGAATAAATCTTCTGTATCCTCGCTCCGTTCCGCCTT

>DEX0356_033.nt.5
CACCGTGGTGTTGGCAGCCTCCTCCGTGTTCATCTCGATGAAGGCCTGGTTTTTCCCCTTCAGCATCAGG
AGGTTGGTGACCTTCCCAAAGGGCAGCCCCAGGGAGATGACTTCCCCCTCCGTGACGTCGATGGGGCTAT
TCCGGCGCCTCCACTCCGTCCCCCGCGGGTCTGCTCTGTGTGCCATGGACGGCATTGTCCCAGATATAGC

CGTTGGTACAAAGCGGGGATCTGACGAGCTTTTCTCTACTTGTGTCACTAACGGACCGTTTATCATGAGC
AGCAACTCGGCTTCTGCAGCAAACGGAAATGACAGCAAGAAGTTCAAAGGTGACAGCCGAAGTGCAGGCG
TCCCCTCTAGAGTGATCCACATCCGGAAGCTCCCCATCGACGTCACGGAGGGGGAAGTCATCTCCCTGGG
GCTGCCCTTTGGGAAGGTCACCAACCTCCTGATGCTGAAGGGGAAAAACCAGGTACCTGAGCCGCGTTTC
TCCGGGGTGCTCACACCGTGCAGGCGGGGACGAGGAGGGCCCAGCGCTCACTGCCTCCCCAACAGGCCTT
CATCGAGATGAACACGGAGGAGGCTGCCAACACCATGGTGAACTACTACACCTCGGTGACCCCTGTGCTG
CGCGGCCAGCCCATCTACATCCAGTTCTCCAACCACAAGGAGCTGAAGACCGACAGCTCTCCCAACCAGG
CGCGGGCCCAGGCGGCCCTGCAGGCGGTGAACTCGGTCCAGTCGGGGAACCTGGCCTTGGCTGCCTCGGC
GGCGGCCGTGGACGCAGGGATGGCGATGGCCGGGCAGAGCCCCGTGCTCAGGATCATCGTGGAGAACCTC
TTCTACCCTGTGACCCTGGATGTGCTGCACCAGATTTTCTCCAAGTTCGGCACAGTGTTGAAGATCATCA
CCTTCACCAAGAACAACCAGTTCCAGGCCCTGCTGCAGTATGCGGACCCCGTGAGCGCCCAGCACGCCAA
GCTGTCGCTGGACGGGCAGAACATCTACAACGCCTGCTGCACGCTGCGCATCGACTTTTCCAAGCTCACC
AGCCTCAACGTCAAGTACAACAATGACAAGAGCCGTGACTACACACGCCCAGACCTGCCTTCCGGGGACA
GCCAGCCCTCGCTGGACCAGACCATGGCCGCGGCCTTCGCCTCTCCGTATGCAGGAGCTGGTTTCCCTCC
CACCTTTGCCATTCCTCAAGCTGCAGGGTATTCAAACGCTTGGTCTTGGTTCCCCAGCGACTGCATGCCC
ACACCACCTTCCCAGGCAGCTCCGCATCCACGGCGGCAGCCTGGGCGGACTGGGCACTCGAGTGCCAGGT
CAGGGGCCCTTCCCGGGAGAGGGGACGCCACGTCCACAGAGCAGGCTTGGCCAGGGCAGTGGTAGGACAG
GGCTGTGGAGGCCCACGTGTGGACGGCGCCAGCCAGAAGCCGCTAATCGCACAGTCTTTGTGGTCTGGTT
CCCTTCAAGCACCTCCGGGCGAGCGCGAGGACCGCCAGTGGTTGGAGGGATGTCTCTAGTAGTTGAATTT
GAGTGGCCTGGTAAGCGCGCGGCCCGAGGCTGAGCAGCATGACCGGCGGCGGGACGTGTGTGCGTGGCCG
TCCTCCCGCTTCCCTTCTGGGACCGGGGCCCGGCCGCCCCAGTGCTGGCCGTGGACGGAGCAGCGCGTGC
CGCCCGGCGGCCGCCTCGTCTGCATGGAGGCATGGGCGCGCATGCGCGGTGGTCCCGGGACGGGCCCTGC
TTGCCGAGGGCCCCGTGTCTGTGCTGGCGGAGCCGGAGCTTTTCTGTGCTGTGGGCGGGGCGCATGCAGA
TGAGCCCAGGCCCGGCCCGGCCCGTGCTGTGAGGAGAGGCGGGCGCTGGTGCAGGAGCCGGCGGGTGGCT
GTGCGGGGGTCGGGCGACCCCACAGGCACCCAGGGTAGGGCCAGAGCCAGGGCCGCCTCCCGCGCGGCTC
GGCTCCTCGGTGCATGAGGACGGGGAGCGTCGGCCTCTCCCACTCTGCGGTGGAGTCGGGGGCGCCGCCG
CTCATCTCACCTCCTGCTTTTCCAGGCCTTTCCGTTCCGAACGTCCACGGCGCCCTGGCCCCCCTGGCCA
TCCCCTCGGCGGCGGCGGCAGCTGCGGCGGCAGGTCGGATCGCCATCCCGGGCCTGGCGGGGGCAGGAAA
TTCTGTATTGCTGGTCAGCAACCTCAACCCAGAGGTACGTGGGCTTTTCCTCCGCGCCGCCGTTCCTCCC
GGAAGAGCGAGCAGGGGATGTTGTGCTCGGGCTCGGGTCCCGGAGCAGCGCCGCCCCTCGCTGTGTCTGC
GCCTCTGCCCCCGGCAGCCCCTCTGCTGCAGCGCTGAGACCCTCCTTTTCCAAGATGGCTTGAGTTTTCCT
CTTTTCCTGAATTCACATCTTGGTTCGCGTTTTCTCTTGCATGATACGCAGAATGAATTATTTTTTGGTT
ACCCAAAGGCAATCTAAAAATTACTTGGAGCTGAGCAGGGCGCGCAGTTCAGAGCCCCAGCAGGCCGCCC
TGCTGGTCAGCGGGACACACAGGGTTAGCGCGGCGGGGTGTGGGCAGCCCCTGTGGCTCGAGAGCCTGTC
ACTTAACTGGTGGATCCGTTATTTCCCAGAATGGTTAAATTCCCCTCGGGGAGAGCTTGGGTGAAACAAG
AGGCCCAAAATATGTAACTCTTGATGGATGGGATCCTCCTGGAATCAGCATTTCAGGCTGTTCTGTTTCC
CGGCAACTCCTGTGACTTCCTCAAGGCGGTGCCCAGCCTGGCCCTCCCTCGCGGGGGCCCAGGACGCACT
CCCTGCCTGTGCCTCTTACCACGAGCCGGGTGCTGTCCCTCCAGCTCCTCCTGAGCGAGGCCGTGCACTA
ACAGACTCGATCTCCCTTCACGCCTGTAATCGTCCTTAGAGACTGACTTGTGCCACAGCTGCGGGAGACA
GCGCTGCTGCAGGGCTTGGCTGAGGTGGCCGCGAGCGGAGCTGCACGCGCGCTCCTGGGGGGCGTGGTTG
GGATCGGCTTCTGGGAGCCCCTCCAGCAGCGCTGTCTCTGCCGGCATGTGAGTGAAGGTGCTTCTCTTAC
CGTGTGCCCGTCGAGGTAAACCTGTAACTGGAATGTGTGTGGAGTGTGACTGATAGAACACTACCTGATT
CTTATGTATTTACTGACCTGTGTTTTTTTGCTACTTTTTTTCTTTTCTCCCCTTCCCCTTTCCCTATTTT
TTTTCTTGCCCTGATCCGGAATTTCTTTGCCAACTGACTGCACGGTACTTCTGCTTCCTGTTGTTGCTTG
AAACAAAACAAAAACATAAACAAATAAAAAACAAAAATTCCCCCTCAAACCCTGCTCTCCGGAAACCAAC
CTGCCCTTGAATATTAACATCCTGACAACTTCATCATCCATCAACCACTGCACGCCTGCGGGGACTGTCT
TCCTCGTGTGGACGATTGGCAACTCGCCCCCCTTGACCTCTCCCTCTCCCCTGTCCCTCCGCTGCCTTGC
TCTGCTGTCTCTAAAGAGAGTCACACCCCAAAGCCTCTTTATTCTTTTCGGTATGTTATCGTTCACACTT
TTATTACCTTGTTTTCATTAATTGAGATGATTTTGATGCCCTGAGACGTATTATGAGTTTGCGGTTTGGC
ACTCTGATGCTCCGTGGCATCCGCCTCGTTTTATGGTTTGCTTTCGGTTTGCGAATTTTATTTGGTCCCG
TAGATACGTACGCATGGTTTATCGCCCTGCATGCTTTTCAGAGTTAGACCTGTCGGTGGCATATGCCACC
GTGGCCACCCGCTGGCAGCTTACCTGTCCTGGATGCTATGACTTTGCTGAACGGAGCTGCTCCTGTTAGC
GCGCCCTGTGGCTGCGAGACGCAGCTCCGCAGTGGCCGATAAAGCAAACCCGGCCGGGCTGAGCCGGGCC
TTGTGGGGGTGCGCGGGGCCGGGGCTGACGGGGAGATGGGCGGGGCAGGCAGCAGGAGACTCAGGCCCCA
TCCCTGGGCTTTTGAAGGCGCGTCTACGGTGACGTGCAGCGCGTGAAGATCCTGTTCAATAAGAAGGAGA
ACGCCCTAGTGCAGATGGCGGACGGCAACCAGGCCCAGCTGGCCATGAGCCACCTGAACGGGCACAAGCT
GCACGGGAAGCCCATCCGCATCACGCTCTCGAAGCACCAGAACGTGCAGCTGCCCCGCGAGGGCCAGGAG
GACCAGGGCCTGACCAAGGACTACGGCAACTCACCCCTGCACCGCTTCAAGAAGCCGGGCTCCAAGAACT
TCCAGAACATATTCCCGCCCTCGGCCACGCTGCACCTCTCCAACATCCCGCCCTCAGTCTCCGAGGAGGA
TCTCAAGGTCCTGTTTTCCAGCAATGGGGGCGTCGTCAAAGGATTCAAGTTCTTCCAGAAGGACCGCAAG

ATGGCACTGATCCAGATGGGCTCCGTGGAGGAGGCGGTCCAGGCCCTCATTGACCTGCACAACCACGACC
TCGGGGAGAACCACCACCTGCGGGTCTCCTTGTCCAAGTCCACCATCTAGGGGCACAGGCCCCCACGGCC
GGGCCCCCTGGCGACAACTTCCATCATTCCAGAGAAAAGCCACTTTAAAAACAGCTGAAGTGACCTTAGC
AGACCAGAGATTTTATTTTTTTAAAGAGAAATCAGTTTACCTGTTTTTAAAAAAATTAAATCTAGTTCAC
CTTGCTCACCCTGCGGTGACAGGGACAGCTCAGGCTCTTGGTGACTGTGGCAGCGGGAGTTCCCGGCCCT
CCACACCCGGGGCCAGACCCTCGGGGCCATGCCTTGGTGGGGCCTGTGTCGGGCGTGGGGCCTGCAGGTG
GGCGCCCCGACCACGACTTGGCTTCCTTGTGCCTTAAAAAACCTGCCTTCCTGCAGCCACACACCCACCC
GGGGTGTCCTGGGGACCCAAGGGGTGGGGGGGTCACACCAGAGAGAGGCAGGGGGCCTGGCCGGCTCCTG
CAGGATCATGCAGCTGGGGCGCGGCGGCCGCGGCTGCGACACCCCAACCCCAGCCCTCTAATCAAGTCAC
GTGATTCTCCCTTCACCCCGCCCCCAGGGCCTTCCCTTCTGCCCCCAGGCGGGCTCCCCGCTGCTCCAGC
TGCGGAGCTGGTCGACATAATCTCTGTATTATATACTTTGCAGTTGCAGACGTCTGTGCCTAGCAATATT
TCCAGTTGACCAAATATTCTAATCTTTTTTCATTTATATGCAAAAGAAATAGTTTTAAGTAACTTTTTAT
AGCAAGATGATACAATGGTATGAGTGTAATCTAAACTTCCTTGTGGTATTACCTTGTATGCTGTTACTTT
TATTTTATTCCTTGTAATTAAGTCACAGGCAGGACCCAGTTTCCAGAGAGCAGGCGGGGCCGCCCAGTGG
GTCAGGCACAGGGAGCCCCGGTCCTATCTTAGAGCCCCTGAGCTTCAGGGAAGGGGCGGGCGTGTCGCCG
CCTCTGGCATCGCCTCCGGTTGCCTTACACCACGCCTTCACCTGCAGTCGCCTAGAAAACTTGCTCTCAA
ACTTCAGGGTTTTTTCTTCCTTCAAATTTTGGACCAAAGTCTCATTTCTGTGTTTTGCCTGCCTCTGATG
CTGGGACCCGGAAGGCGGGCGCTCCTCCTGTCTTCTCTGTGCTCTTTCTACCGCCCCCGCGTCCTGTCCC
GGGGGCTCTCCTAGGATCCCCTTTCCGTAAAAGCGTGTAACAAGGGTGTAAATATTTATAATTTTTTATA
CCTGTTGTGAGACCCGAGGGGCGGCGGCGCGGTTTTTTATGGTGACACAAATGTATATTTTGCTAACAGC
AATTCCAGGCTCAGTATTGTGACCGCGGAGCCACAGGGGACCCCACGCACATTCCGTTGCCTTACCCGAT
GGCTTGTGACGCGGAGAGAACCGATTAAAACCGTTTGAGAAACTCCTCCCTTGTCTAGCCCTGTGTTCGC
TGTGGACGCTGTAGAGGCAGGTTGGCCAGTCTGTACCTGGACTTCGAATAAATCTTCTGTATCCTCGCTC
CGTTCCGCCTTCGTTGCTGGCTTCCTCTTTCTGTGCCGCGGGCGGCCCCGACTCCCGCCCGGGTGTCCGC
CGAAGCGCCAGGTGGGGCCACGGGCGGGGTGGGCCCGGCCCTGAGAGTGGGGGTCCCTGCGTGGGGCGTG
CGGGCCAGAGCGAATGACCAAGAGCCGGACCTCGGTTTCTGCCGCTTTATTGAGCATCCGCGCGGCCGC

>DEX0356_033.nt.6

CACCGTGGTGTTGGCAGCCTCCTCCGTGTTCATCTCGATGAAGGCCTGGTTTTTCCCCTTCAGCATCAGG
AGGTTGGTGACCTTCCCAAAGGGCAGCCCCAGGGAGATGACTTCCCCCTCCGTGACGTCGATGGGGCTAT
TCCGGCGCCTCCACTCCGTCCCCCGCGGGTCTGCTCTGTGTGCCATGGACGGCATTGTCCCAGATATAGC
CGTTGGTACAAAGCGGGGATCTGACGAGCTTTTCTCTACTTGTGTCACTAACGGACCGTTTATCATGAGC
AGCAACTCGGCTTCTGCAGCAAACGGAAATGACAGCAAGAAGTTCAAAGGTGACAGCCGAAGTGCAGGCG
TCCCCTCTAGAGTGATCCACATCCGGAAGCTCCCCATCGACGTCACGGAGGGGGAAGTCATCTCCCTGGG
GCTGCCCTTTGGGAAGGTCACCAACCTCCTGATGCTGAAGGGGAAAAACCAGGTACCTGAGCCGCGTTTC
TCCGGGGTGCTCACACCGTGCAGGCGGGGACGAGGAGGGCCCAGCGCTCACTGCCTCCCCAACAGGCCTT
CATCGAGATGAACACGGAGGAGGCTGCCAACACCATGGTGAACTACTACACCTCGGTGACCCCTGTGCTG
CGCGGCCAGCCCATCTACATCCAGTTCTCCAACCACAAGGAGCTGAAGACCGACAGCTCTCCCCAACCAGG
CGCGGGCCCAGGCGGCCCTGCAGGCGGTGAACTCGGTCCAGTCGGGGAACCTGGCCTTGGCTGCCTCGGC
GGCGGCCGTGGACGCAGGGATGGCGATGGCCGGGCAGAGCCCCGTGCTCAGGATCATCGTGGAGAACCTC
TTCTACCCTGTGACCCTGGATGTGCTGCACCAGATTTTCTCCAAGTTCGGCACAGTGTTGAAGATCATCA
CCTTCACCAAGAACAACCAGTTCCAGGCCCTGCTGCAGTATGCGGACCCCGTGAGCGCCCAGCACGCCAA
GCTGTCGCTGGACGGGCAGAACATCTACAACGCCTGCTGCACGCTGCGCATCGACTTTTCCAAGCTCACC
AGCCTCAACGTCAAGTACAACAATGACAAGAGCCGTGACTACACACGCCCAGACCTGCCTTCCGGGGACA
GCCAGCCCTCGCTGGACCAGACCATGGCCGCGGCCTTCGCCTCTCCGTATGCAGGAGCTGGTTTCCCTCC
CACCTTTGCCATTCCTCAAGCTGCAGGGTATTCAAACGCTTGGTCTTGGTTCCCCAGCGACTGCATGCCC
ACACCACCTTCCCAGGCAGCTCCGCATCCACGGCGGCAGCCTGGGCGGACTGGGCACTCGAGTGCCAGGT
CAGGGGCCCTTCCCGGGAGAGGGGACGCCACGTCCACAGAGCAGGCTTGGCCAGGGCAGTGGTAGGACAG
GGCTGTGGAGGCCCACGTGTGGACGGCGCCAGCCAGAAGCCGCTAATCGCACAGTCTTTGTGGTCTGGTT
CCCTTCAAGCACCTCCGGGCGAGCGCGAGGACCGCCAGTGGTTGGAGGGATGTCTCTAGTAGTTGAATTT
GAGTGGCCTGGTAAGCGCGCGGCCCGAGGCTGAGCAGCATGACCGGCGGCGGGACGTGTGTGCGTGGCCG
TCCTCCCGCTTCCCTTCTGGGACCGGGGCCCGGCCGCCCCAGTGCTGGCCGTGGACGGAGCAGCGCGTGC
CGCCCGGCGGCCGCCTCGTCTGCATGGAGGCATGGGCGCGCATGCGCGGTGGTCCCGGGACGGGCCCTGC
TTGCCGAGGGCCCCGTGTCTGTGCTGGCGGAGCCGGAGCTTTTCTGTGCTGTGGGCGGGGCGCATGCAGA
TGAGCCCAGGCCCGGCCCGGCCCGTGCTGTGAGGAGAGGCGGGCGCTGGTGCAGGAGCCGGCGGGTGGCT
GTGCGGGGGTCGGGCGACCCCACAGGCACCCAGGGTAGGGCCAGAGCCAGGGCCGCCTCCCGCGCGGCTC
GGCTCCTCGGTGCATGAGGACGGGGAGCGTCGGCCTCTCCCACTCTGCGGTGGAGTCGGGGGCGCCGCCG
CTCATCTCACCTCCTGCTTTTCCAGGCCTTTCCGTTCCGAACGTCCACGGCGCCCTGGCCCCCCTGGCCA
TCCCCTCGGCGGCGGCGGCAGCTGCGGCGGCAGGTCGGATCGCCATCCCGGGCCTGGCGGGGGCAGGAAA
TTCTGTATTGCTGGTCAGCAACCTCAACCCAGAGGTACGTGGGCTTTTCCTCCGCGCCGCCGTTCCTCCC

GGAAGAGCGAGCAGGGGATGTTGTGCTCGGGCTCGGGTCCCGGAGCAGCGCCGCCCCTCGCTGTGTCTGC
GCCTCTGCCCCGGCAGCCCCTCTGCTGCAGCGCTGAGACCCTCCTTTTCCAAGATGGCTTGAGTTTTCCT
CTTTTCCTGAATTCACATCTTGGTTCGCGTTTTCTCTTGCATGATACGCAGAATGAATTATTTTTTGGTT
ACCCAAAGGCAATCTAAAAATTACTTGGAGCTGAGCAGGGCGCGCAGTTCAGAGCCCCAGCAGGCCGCCC
TGCTGGTCAGCGGGACACACAGGGTTAGCGCGGCGGGGTGTGGGCAGCCCCTGTGGCTCGAGAGCCTGTC
ACTTAACTGGTGGATCCGTTATTTCCCAGAATGGTTAAATTCCCCTCGGGGAGAGCTTGGGTGAAACAAG
AGGCCCAAAATATGTAACTCTTGATGGATGGGATCCTCCTGGAATCAGCATTTCAGGCTGTTCTGTTTCC
CGGCAACTCCTGTGACTTCCTCAAGGCGGTGCCCAGCCTGGCCCTCCCTCGCGGGGGCCCAGGACGCACT
CCCTGCCTGTGCCTCTTACCACGAGCCGGGTGCTGTCCCTCCAGCTCCTCCTGAGCGAGGCCGTGCACTA
ACAGACTCGATCTCCCTTCACGCCTGTAATCGTCCTTAGAGACTGACTTGTGCCACAGCTGCGGGAGACA
GCGCTGCTGCAGGGCTTGGCTGAGGTGGCCGCGAGCGGAGCTGCACGCGCGCTCCTGGGGGGCGTGGTTG
GGATCGGCTTCTGGGAGCCCCTCCAGCAGCGCTGTCTCTGCCGGCATGTGAGTGAAGGTGCTTCTCTTAC
CGTGTGCCCGTCGAGGTAAACCTGTAACTGGAATGTGTGTGGAGTGTGACTGATAGAACACTACCTGATT
CTTATGTATTTACTGACCTGTGTTTTTTTGCTACTTTTTTTCTTTTCTCCCCTTCCCCTTTCCCTATTTT
TTTTCTTGCCCTGATCCGGAATTTCTTTGCCAACTGACTGCACGGTACTTCTGCTTCCTGTTGTTGCTTG
AAACAAAACAAAAACATAAACAAATAAAAAACAAAAATTCCCCCTCAAACCCTGCTCTCCGGAAACCAAC
CTGCCCTTGAATATTAACATCCTGACAACTTCATCATCCATCAACCACTGCACGCCTGCGGGGACTGTCT
TCCTCGTGTGGACGATTGGCAACTCGCCCCCCTTGACCTCTCCCTCTCCCCTGTCCCTCCGCTGCCTTGC
TCTGCTGTCTCTAAAGAGAGTCACACCCCAAAGCCTCTTTATTCTTTTCGGTATGTTATCGTTCACACTT
TTATTACCTTGTTTTCATTAATTGAGATGATTTTGATGCCCTGAGACGTATTATGAGTTTGCGGTTTGGC
ACTCTGATGCTCCGTGGCATCCGCCTCGTTTTATGGTTTGCTTTCGGTTTGCGAATTTTATTTGGTCCCG
TAGATACGTACGCATGGTTTATCGCCCTGCATGCTTTTCAGAGTTAGACCTGTCGGTGGCATATGCCACC
GTGGCCACCCGCTGGCAGCTTACCTGTCCTGGATGCTATGACTTTGCTGAACGGAGCTGCTCCTGTTAGC
GCGCCCTGTGGCTGCGAGACGCAGCTCCGCAGTGGCCGATAAAGCAAACCCGGCCGGGCTGAGCCGGGCC
TTGTGGGGGTGCGCGGGGCCGGGGCTGACGGGGAGATGGGCGGGGCAGGCAGCAGGAGACTCAGGCCCCA
TCCCTGGGCTTTTGAAGGCGCGTCTACGGTGACGTGCAGCGCGTGAAGATCCTGTTCAATAAGAAGGAGA
ACGCCCTAGTGCAGATGGCGGACGGCAACCAGGCCCAGCTGGCCATGAGCCACCTGAACGGGCACAAGCT
GCACGGGAAGCCCATCCGCATCACGCTCTCGAAGCACCAGAACGTGCAGCTGCCCCGCGAGGGCCAGGAG
GACCAGGGCCTGACCAAGGACTACGGCAACTCACCCCTGCACCGCTTCAAGAAGCCGGGCTCCAAGAACT
TCCAGAACATATTCCCGCCCTCGGCCACGCTGCACCTCTCCAACATCCCGCCCTCAGTCTCCGAGGAGGA
TCTCAAGGTCCTGTTTTCCAGCAATGGGGGCGTCGTCAAAGGATTCAAGTTCTTCCAGAAGGACCGCAAG
ATGGCACTGATCCAGATGGGCTCCGTGGAGGAGGCGGTCCAGGCCCTCATTGACCTGCACAACCACGACC
TCGGGGAGAACCACCACCTGCGGGTCTCCTTCTCCAAGTCCACCATCTAGGGGCACAGGCCCCCACGGCC
GGGCCCCCTGGCGACAACTTCCATCATTCCAGAGAAAAGCCACTTTAAAAACAGCTGAAGTGACCTTAGC
AGACCAGAGATTTTATTTTTTTAAAGAGAAATCAGTTTACCTGTTTTTAAAAAAATTAAATCTAGTTCAC
CTTGCTCACCCTGCGGTGACAGGGACAGCTCAGGCTCTTGGTGACTGTGGCAGCGGGAGTTCCCGGCCCT
CCACACCCGGGGCCAGACCCTCGGGGCCATGCCTTGGTGGGGCCTGTGTCGGGCGTGGGGCCTGCAGGTG
GGCGCCCCGACCACGACTTGGCTTCCTTGTGCCTTAAAAAACCTGCCTTCCTGCAGCCACACACCCACCC
GGGGTGTCCTGGGGACCCAAGGGGTGGGGGGGTCACACCAGAGAGAGGCAGGGGGCCTGGCCGGCTCCTG
CAGGATCATGCAGCTGGGGCGCGGCGGCCGCGGCTGCGACACCCCAACCCCAGCCCTCTAATCAAGTCAC
GTGATTCTCCCTTCACCCCGCCCCCAGGGCCTTCCCTTCTGCCCCCAGGCGGGCTCCCCGCTGCTCCAGC
TGCGGAGCTGGTCGACATAATCTCTGTATTATATACTTTGCAGTTGCAGACGTCTGTGCCTAGCAATATT
TCCAGTTGACCAAATATTCTAATCTTTTTTCATTTATATGCAAAAGAAATAGTTTTAAGTAACTTTTTAT
AGCAAGATGATACAATGGTATGAGTGTAATCTAAACTTCCTTGTGGTATTACCTTGTATGCTGTTACTTT
TATTTTATTCCTTGTAATTAAGTCACAGGCAGGACCCAGTTTCCAGAGAGCAGGCGGGGCCGCCCAGTGG
GTCAGGCACAGGGAGCCCCGGTCCTATCTTAGAGCCCCTGAGCTTCAGGGAAGGGGCGGGCGTTCTCTGT
GCTCTTTCTACCGCCCCCGCGTCCTGTCCCGGGGGCTCTCCTAGGATCCCCTTTCCGTAAAAGCGTGTAA
CAAGGGTGTAAATATTTATAATTTTTTATACCTGTTGTGAGACCCGAGGGGCGGCGGCGCGGTTTTTTAT
GGTGACACAAATGTATATTTTGCTAACAGCAATTCCAGGCTCAGTATTGTGACCGCGGAGCCACAGGGGA
CCCCACGCACATTCCGTTGCCTTACCCGATGGCTTGTGACGCGGAGAGAACCGATTAAAACCGTTTGAGA
AACTC

>DEX0356_033.nt.7
CACCGTGGTGTTGGCAGCCTCCTCCGTGTTCATCTCGATGAAGGCCTGGTTTTTCCCCTTCAGCATCAGG
AGGTTGGTGACCTTCCCAAAGGGCAGCCCCAGGGAGATGACTTCCCCCTCCGTGACGTCGATGGGGCTAT
TCCGGCGCCTCCACTCCGTCCCCCGCGGGTCTGCTCTGTGTGCCATGGACGGCATTGTCCCAGATATAGC
CGTTGGTACAAAGCGGGGATCTGACGAGCTTTTCTCTACTTGTGTCACTAACGGACCGTTTATCATGAGC
AGCAACTCGGCTTCTGCAGCAAACGGAAATGACAGCAAGAAGTTCAAAGGTGACAGCCGAAGTGCAGGCG
TCCCCTCTAGAGTGATCCACATCCGGAAGCTCCCCATCGACGTCACGGAGGGGGAAGTCATCTCCCTGGG
GCTGCCCTTTGGGAAGGTCACCAACCTCCTGATGCTGAAGGGGAAAAACCAGGTACCTGAGCCGCGTTTC

TCCGGGGTGCTCACACCGTGCAGGCGGGGACGAGGAGGGCCCAGCGCTCACTGCCTCCCCAACAGGCCTT
CATCGAGATGAACACGGAGGAGGCTGCCAACACCATGGTGAACTACTACACCTCGGTGACCCCTGTGCTG
CGCGGCCAGCCCATCTACATCCAGTTCTCCAACCACAAGGAGCTGAAGACCGACAGCTCTCCCAACCAGG
CGCGGGCCCAGGCGGCCCTGCAGGCGGTGAACTCGGTCCAGTCGGGGAACCTGGCCTTGGCTGCCTCGGC
GGCGGCCGTGGACGCAGGGATGGCGATGGCCGGGCAGAGCCCCGTGCTCAGGATCATCGTGGAGAACCTC
TTCTACCCTGTGACCCTGGATGTGCTGCACCAGATTTTCTCCAAGTTCGGCACAGTGTTGAAGATCATCA
CCTTCACCAAGAACAACCAGTTCCAGGCCCTGCTGCAGTATGCGGACCCCGTGAGCGCCCAGCACGCCAA
GCTGTCGCTGGACGGGCAGAACATCTACAACGCCTGCTGCACGCTGCGCATCGACTTTTCCAAGCTCACC
AGCCTCAACGTCAAGTACAACAATGACAAGAGCCGTGACTACACACGCCCAGACCTGCCTTCCGGGGACA
GCCAGCCCTCGCTGGACCAGACCATGGCCGCGGCCTTCGCCTCTCCGTATGCAGGAGCTGGTTTCCCTCC
CACCTTTGCCATTCCTCAAGCTGCAGGGTATTCAAACGCTTGGTCTTGGTTCCCCAGCGACTGCATGCCC
ACACCACCTTCCCAGGCAGCTCCGCATCCACGGCGGCAGCCTGGGCGGACTGGGCACTCGAGTGCCAGGT
CAGGGGCCCTTCCCGGGAGAGGGGACGCCACGTCCACAGAGCAGGCTTGGCCAGGGCAGTGGTAGGACAG
GGCTGTGGAGGCCCACGTGTGGACGGCGCCAGCCAGAAGCCGCTAATCGCACAGTCTTTGTGGTCTGGTT
CCCTTCAAGCACCTCCGGGCGAGCGCGAGGACCGCCAGTGGTTGGAGGGATGTCTCTAGTAGTTGAATTT
GAGTGGCCTGGTAAGCGCGCGGCCCGAGGCTGAGCAGCATGACCGGCGGCGGGACGTGTGTGCGTGGCCG
TCCTCCCGCTTCCCTTCTGGGACCGGGGCCCGGCCGCCCCAGTGCTGGCCGTGGACGGAGCAGCGCGTGC
CGCCCGGCGGCCGCCTCGTCTGCATGGAGGCATGGGCGCGCATGCGCGGTGGTCCCGGGACGGGCCCTGC
TTGCCGAGGGCCCCGTGTCTGTGCTGGCGGAGCCGGAGCTTTTCTGTGCTGTGGGCGGGGCGCATGCAGA
TGAGCCCAGGCCCGGCCCGGCCCGTGCTGTGAGGAGAGGCGGGCGCTGGTGCAGGAGCCGGCGGGTGGCT
GTGCGGGGGTCGGGCGACCCCACAGGCACCCAGGGTAGGGCCAGAGCCAGGGCCGCCTCCCGCGCGGCTC
GGCTCCTCGGTGCATGAGGACGGGGAGCGTCGGCCTCTCCCACTCTGCGGTGGAGTCGGGGGCGCCGCCG
CTCATCTCACCTCCTGCTTTTCCAGGCCTTTCCGTTCCGAACGTCCACGGCGCCCTGGCCCCCCTGGCCA
TCCCCTCGGCGGCGGCGGCAGCTGCGGCGGCAGGTCGGATCGCCATCCCGGGCCTGGCGGGGGCAGGAAA
TTCTGTATTGCTGGTCAGCAACCTCAACCCAGAGGTACGTGGGCTTTTCCTCCGCGCCGCCGTTCCTCCC
GGAAGAGCGAGCAGGGGATGTTGTGCTCGGGCTCGGGTCCCGGAGCAGCGCCGCCCCTCGCTGTGTCTGC
GCCTCTGCCCCGGCAGCCCCTCTGCTGCAGCGCTGAGACCCTCCTTTTCCAAGATGGCTTGAGTTTTCCT
CTTTTCCTGAATTCACATCTTGGTTCGCGTTTTCTCTTGCATGATACGCAGAATGAATTATTTTTTGGTT
ACCCAAAGGCAATCTAAAAATTACTTGGAGCTGAGCAGGGCGCGCAGTTCAGAGCCCCAGCAGGCCGCCC
TGCTGGTCAGCGGGACACACAGGGTTAGCGCGGCGGGGTGTGGGCAGCCCCTGTGGCTCGAGAGCCTGTC
ACTTAACTGGTGGATCCGTTATTTCCCAGAATGGTTAAATTCCCCTCGGGGAGAGCTTGGGTGAAACAAG
AGGCCCAAAATATGTAACTCTTGATGGATGGGATCCTCCTGGAATCAGCATTTCAGGCTGTTCTGTTTCC
CGGCAACTCCTGTGACTTCCTCAAGGCGGTGCCCAGCCTGGCCCTCCCTCGCGGGGGCCCAGGACGCACT
CCCTGCCTGTGCCTCTTACCACGAGCCGGGTGCTGTCCCTCCAGCTCCTCCTGAGCGAGGCCGTGCACTA
ACAGACTCGATCTCCCTTCACGCCTGTAATCGTCCTTAGAGACTGACTTGTGCCACAGCTGCGGGAGACA
GCGCTGCTGCAGGGCTTGGCTGAGGTGGCCGCGAGCGGAGCTGCACGCGCGCTCCTGGGGGGCGTGGTTG
GGATCGGCTTCTGGGAGCCCCTCCAGCAGCGCTGTCTCTGCCGGCATGTGAGTGAAGGTGCTTCTCTTAC
CGTGTGCCCGTCGAGGTAAACCTGTAACTGGAATGTGTGTGGAGTGTGACTGATAGAACACTACCTGATT
CTTATGTATTTACTGACCTGTGTTTTTTTGCTACTTTTTTTCTTTTCTCCCCTTCCCCTTTCCCTATTTT
TTTTCTTGCCCTGATCCGGAATTTCTTTGCCAACTGACTGCACGGTACTTCTGCTTCCTGTTGTTGCTTG
AAACAAAACAAAAACATAAACAAATAAAAAACAAAAATTCCCCCTCAAACCCTGCTCTCCGGAAACCAAC
CTGCCCTTGAATATTAACATCCTGACAACTTCATCATCCATCAACCACTGCACGCCTGCGGGGACTGTCT
TCCTCGTGTGGACGATTGGCAACTCGCCCCCCTTGACCTCTCCCTCTCCCCTGTCCCTCCGCTGCCTTGC
TCTGCTGTCTCTAAAGAGAGTCACACCCCAAAGCCTCTTTATTCTTTTCGGTATGTTATCGTTCACACTT
TTATTACCTTGTTTTCATTAATTGAGATGATTTTGATGCCCTGAGACGTATTATGAGTTTGCGGTTTGGC
ACTCTGATGCTCCGTGGCATCCGCCTCGTTTTATGGTTTGCTTTCGGTTTGCGAATTTTATTTGGTCCCG
TAGATACGTACGCATGGTTTATCGCCCTGCATGCTTTTCAGAGTTAGACCTGTCGGTGGCATATGCCACC
GTGGCCACCCGCTGGCAGCTTACCTGTCCTGGATGCTATGACTTTGCTGAACGGAGCTGCTCCTGTTAGC
GCGCCCTGTGGCTGCGAGACGCAGCTCCGCAGTGGCCGATAAAGCAAACCCGGCCGGGCTGAGCCGGGCC
TTGTGGGGGTGCGCGGGGCCGGGGCTGACGGGGAGATGGGCGGGGCAGGCAGCAGGAGACTCAGGCCCCA
TCCCTGGGCTTTTGAAGGCGCGTCTACGGTGACGTGCAGCGCGTGAAGATCCTGTTCAATAAGAAGGAGA
ACGCCCTAGTGCAGATGGCGGACGGCAACCAGGCCCAGCTGGCCATGAGCCACCTGAACGGGCACAAGCT
GCACGGGAAGCCCATCCGCATCACGCTCTCGAAGCACCAGAACGTGCAGCTGCCCCGCGAGGGCCAGGAG
GACCAGGGCCTGACCAAGGACTACGGCAACTCACCCCTGCACCGCTTCAAGAAGCCGGGCTCCAAGAACT
TCCAGAACATATTCCCGCCCTCCCCACGGCCGGGCCCCCTGGCGACAACTTCCATCATTCCAGAGAAAAG
CCACTTTAAAAACAGCTGAAGTGACCTTAGCAGACCAGAGATTTTATTTTTTTAAAGAGAAATCAGTTTA
CCTGTTTTTAAAAAAATTAAATCTAGTTCACCTTGCTCACCCTGCGGTGACAGGGACAGCTCAGGGCCGG
CTCCTGCAGGATCATGCAGCTGGGGCGCGGCGGCCGCGGCTGCGACACCCCAACCCCAGCCCTCTAATCA
AGTCACGTGATTCTCCCTTCACCCCGCCCCCAGGGCCTTCCCTTCTGCCCCCAGGCGGGCTCCCCGCTGC
TCCAGCTGCGGAGCTGGTCGACATAATCTCTGTATTATATACTTTGCAGTTGCAGACGTCTGTGCCTAGC

AATATTTCCAGTTGACCAAATATTCTAATCTTTTTTCATTTATATGCAAAAGAAATAGTTTTAAGTAACT
TTTTATAGCAAGATGATACAATGGTATGAGTGTAATCTAAACTTCCTTGTGGTATTACCTTGTATGCTGT
TACTTTTATTTTATTCCTTGTAATTAAGTCACAGGCAGGACCCAGTTTCCAGAGAGCAGGCGGGGCCGCC
CAGTGGGTCAGGCACAGGGAGCCCCGGTCCTATCTTAGAGCCCCTGAGCTTCAGGGAAGGGGCGGGCGTG
TCGCCGCCTCTGGCATCGCCTCCGGTTGCCTTACACCACGCCTTCACCTGCAGTCGCCTAGAAAACTTGC
TCTCAAACTTCAGGGTTTTTTCTTCCTTCAAATTTTGGACCAAAGTCTCATTTCTGTGTTTTGCCTGCCT
CTGATGCTGGGACCCGGAAGGCGGGCGCTCCTCCTGTCTTCTCTGTGCTCTTTCTACCGCCCCCGCGTCC
TGTCCCGGGGGCTCTCCTAGGATCCCCTTTCCGTAAAAGCGTGTAACAAGGGTGTAAATATTTATAATTT
TTTATACCTGTTGTGAGACCCGAGGGGCGGCGGCGCGGTTTTTTATGGTGACACAAATGTATATTTTGCT
AACAGCAATTCCAGGCTCAGTATTGTGACCGCGGAGCCACAGGGGACCCCACGCACATTCCGTTGCCTTA
CCCGATGGCTTGTGACGCGGAGAGAACCGATTAAAACCGTTTGAGAAACTCCTCCCTTGTCTAGCCCTGT
GTTCGCTGTGGACGCTGTAGAGGCAGGTTGGCCAGTCTGTACCTGGACTTCGAATAAATCTTCTGTATCC
TCGCTCCGTTCCGCCTT

>DEX0356_034.nt.1
AACATAGGTCAGTACAGGCCGGGTTATCGTCGCAAGAAAGCCTGAGTCCTGTCCTTTCTCTCTCCCCGGA
CAGCATGAGCTTCACCACTCGCTCCACCTTCTCCACCAACTACCGGTCCCTGGGCTCTGTCCAGGCGCCC
AGCTACGGCGCCCGGCCGGTCAGCAGCGCGGCCAGCGTCTATGCAGGCGCTGGGGGCTCTGGTTCCCGGA
TCTCCGTGTCCCGCTCCACCAGCTTCAGGGGCGGCATGGGGTCCGGGGGCCTGGCCACCGGGATAGCCGG
GGGTCTGGCAGGAATGGGAGGCATCCAGAACGAGAAGGAGACCATGCAAAGCCTGAACGACCGCCTGGCC
TCTTACCTGGACAGAGTGAGGAGCCTGGAGACCGAGAACCGGAGGCTGGAGAGCAAAATCCGGGAGCACT
TGGAGAAGAAGGGACCCCAGGTCAGAGACTGGAGCCATTACTTCAAGATCATCGAGGACCTGAGGGCTCA
GATCTTCGCAAATACTGTGGACAATGCCCGCATCGTTCTGCAGATTGACAATGCCCGTCTTGCTGCTGAT
GACTTTAGAGTCAAGTATGAGACAGAGCTGGCCATGCGCCAGTCTGTGGAGAACGACATCCATGGGCTCC
GCAAGGTCATTGATGACACCAATATCACACGACTGCAGCTGGAGACAGAGATCGAGGCTCTCAAGGAGGA
GCTGCTCTTCATGAAGAAGAACCACGAAGAGGACSGTACAGGSCTACAAGMCCMGATGSCMGSTCTGGGT
GMCGYGGAGGAGATGCCCCCAACACCGGGACTGGCAGACCAGGCGACTCGGCAAGAGACGAGTGGTCGAG
ACCGAGGGGAGAAGGTCGCCAGAATGGGGAGCCACGGCCCAAGCGGAGTGTGAACGAGCCGGTAGAGCGG
CCCGGGAGCGACGATAGCGGGCACGGGGGCACAG

>DEX0356_034.nt.2
GAAAGCAGCCTCGAGGGCCAACAACACCTGCTGTCCGTGTCCATGCCCGGTTGGCCACCCCGTTTCTGGG
GGCATGAGCTTCACCACTCGCTCCACCTTCTCCACCAACTACCGGTCCCTGGGCTCTGTCCAGGCGCCCA
GCTACGGCGCCCGGCCGGTCAGCAGCGCGGCCAGCGTCTATGCAGGCGCTGGGGGCTCTGGTTCCCGGAT
CTCCGTGTCCCGCTCCACCAGCTTCAGGGGCGGCATGGGGTCCGGGGGCCTGGCCACCGGGATAGCCGGG
GGTCTGGCAGGAATGGGAGGCATCCAGAACGAGAAGGAGACCATGCAAAGCCTGAACGACCGCCTGGCCT
CTTACCTGGACAGAGTGAGGAGCCTGGAGACCGAGAACCGGAGGCTGGAGAGCAAAATCCGGGAGCACTT
GGAGAAGAAGGGACCCCAGGTCAGAGACTGGAGCCATTACTTCAAGATCATCGAGGACCTGAGGGCTCAG
ATCTTCGCAAATACTGTGGACAATGCCCGCATCGTTCTGCAGATTGACAATGCCCGTCTTGCTGCTGATG
ACTTTAGAGTCAAGTATGAGACAGAGCTGGCCATGCGCCAGTCTGTGGAGAACGACATCCATGGGCTCCG
CAAGGTCATTGATGACACCAATATCACACGACTGCAGCTGGAGACAGAGATCGAGGCTCTCAAGGAGGAG
CTGCTCTTCATGAAGAAGAACCACGAAGAGGAAGTAAAAGGCCTACAAGCCCAGATTGCCAGCTCTGGGT
TGACCGTGGAGGTAGATGCCCCCAAATCTCAGGACCTCGCCAAGATCATGGCAGACATCCGGGCCCAATA
TGACGAGCTGGCTCGGAAGAACCGAGAGGAGCTAGACAAGTACTGGTCTCAGCAGATTGAGGAGAGCACC
ACAGTGGTCACCACACAGTCTGCTGAGGTTGGAGCTGCTGAGACGACGCTCACAGAGCTGAGACGTACAG
TCCAGTCCTTGGAGATCGACCTGGACTCCATGAGAAATCTGAAGGCCAGCTTGGAGAACAGCCTGAGGGA
GGTGGAGGCCCGCTACGCCCTACAGATGGAGCAGCTCAACGGGATCCTGCTGCACCTTGAGTCAGAGCTG
GCACAGACCCGGGCAGAGGGACAGCGCCAGGCCCAGGAGTATGAGGCCCTGCTGAACATCAAGGTCAAGC
TGGAGGCTGAGATCGCCACCTACCGCCGCCTGCTGGAAGATGGCGAGGACTTTAATCTTGGTGATGCCTT
GGACAGCAGCAACTCCATGCAAACCATCCAAAAGACCACCACCCGCCGGATAGTGGATGGCAAAGTGGTG
TCTGAGACCAATGACACCAAAGTTCTGAGGCATTAAGCCAGCAGAAGCAGGGTACCCTTTGGGGAGCAGG
AGGCCAATAAAAAGTTCAGAGTTCAT

>DEX0356_034.nt.3
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGGGGACTGGTCTCAGCAGATTGAGGAGAGCACCACAGTGG
TCACCACACAGTCTGCTGAGGTTGGAGCTGCTGAGACGACGCTCACAGAGCTGAGACGTACAGTCCAGTC
CTTGGAGATCGACCTGGACTCCATGAGAAATCTGAAGGCCAGCTTGGAGAACAGCCTGAGGGAGGTGGAG
GCCCGCTACGCCCTACAGATGGAGCAGCTCAACGGGATCCTGCTGCACCTTGAGTCAGAGCTGGCACAGA
CCCGGGCAGAGGGACAGCGCCAGGCCCAGGAGTATGAGGCCCTGCTGAACATCAAGGTCAAGCTGGAGGC
TGAGATCGCCACCTACCGCCGCCTGCTGGAAGATGGCGAGGACTTTAATCTTGGTGATGCCTTGGACAGC

AGCAACTCCATGCAAACCATCCAAAAGACCACCACCCGCCGGATAGTGGATGGCAAAGTGGTGTCTGAGA
CCAATGACACCAAAGTTCTGAGGCATTAAGCCAGCAGAAGCAGGGTACCCTTTGGGGAGCAGGAGGCCAA
TAAAAAGTTCAGAGTTCAT

>DEX0356_035.nt.1
CCCAAAGAAAGCTTGGCTATTTTTATTGATGTGTAATATACAACCTATGTAAATGAAGTTAGGCCTATTG
GTTTGCAAATGCAGCTTTAACATAATTACCTTACCTGTCTCCTTCCCCTACCCAATGCTGAGGGACATTG
CTCCCCACCTCACCATCATGCCATGCTTTCTCCCCCTGGTCATAGGTGATCTTTCCAGAACAGCTAACCA
GGTGCCTGGGGTCTGGAGACCTTACTGCTTGAGGAGTGAATTAAGAGAAAAGACTGCTTGCTTTCCTCCA
GACTTTGAGCCCTGGCCTGATGTAGACCTTTTTGCTCTCTCCTCCTTCGTATAGGTACGGTTCCTGGAGC
AGCAGAACAAGATGCTGGAGACCAAGTGGAGCCTCCTGCAGCAGCAGAAGACGGCTCGAAGCAACATGGA
CAACATGTTCGAGAGCTACATCAACAACCTTAGGCGGCAGCTGGAGACTCTGGGCCAGGAGAAGCTGAAG
CTGGAGGCGGAGCTTGGCAACATGCAGGGGCTGGTGGAGGACTTCAAGAACAAAAGTGAGCAACTCCCAC
CCTCCACCCAACTGAAGTCACCGTGCTCTCCTCCACCCCTTGGGACCTTGGGACTAAGTCCCCATGGCCC
TCTGTTGTGGGCAGTGCAGTCCTATCTAATTAGGGTGACCACCTGATGAGGTTTGCTCGGAACAGTCTGT
GTTTATGCCCAGGTTCCTAGCACAATTGTTGATAGTACC

>DEX0356_036.nt.1
ATTGGAAGCGCACTGCGCGGCCTCCACCGCGCTCTCGGCCCGGGTTCGGCGTGGGGGCGGGGAGGTAAGT
ACGGTAATGGGCGGGGCCGAGCGCGGGCGCGCTCCGGCTTTCCTCCTGCGCAGTGCTCCCGTCAGCGCAG
GGGGCGGGGCGCCTATATTACGTGCGCGGCGCCGCTCCTGCGAGAGGACGTTGCGTGCTCGCTCGCGCCA
GGCGAGTCTCCGCGTCTCCCTCGCGAACTCGGTGAAAGGAATTGGCGCCGTTCGACACCAGGCGGATCCG
CTCTGCAGCACGAACCCATCTCCAGCCGCAGCCGCAGCCGCCGCCCGGGCCGAGGAGCCGCCTCGGAAGA
CGAGAACGGCGACGATCAGGGTTTCCAGGAAGGGGAAGATGAGCTCGGGGACGAAGAGGAAGGCGCGGGC
GACGAGAACGGGCACGGGGAGCAGCAGCCTCAACCGCCGGCGACGCAGCAGCAACAGCCCCAACAGCAGC
GCGGGGCCGCCAAGGAGGCCGCGGGGAAGAGCAGCGGCCCCACCTCGCTGTTCGCGGTGACGGTGGCGCC
GCCCGGGGCGAGGCAGGGCCAGCAGCAGGCGGGAGGGGACGGCAAAACAGAACAGAAAGGCGGAGATAAA
AAGAGGGGTGTTAAAAGACCACGAGAAGATCATGGCCGTGGATATTTTGAGTACATTGAAGAGAACAAGT
ATAGCAGAGCCAAATCTCCTCAGCCACCTGTTGAAGAAGAAGATGAACACTTCGATGACACAGTGGTTTG
TCTTGATACTTATAATTGTGATCTACATTTTAAAATATCAAGAGATCGTCTCAGTGCTTCTTCCCTTACA
ATGGAGAGTTTTGCTTTTCTTTGGGCTGGAGGAAGAGCATCCTATGGTGTGTCAAAAGGCAAAGTGTGTT
TTGAGATGAAGGTTACAGAGAAGATCCCAGTAAGGCATTTATATACAAAAGATATTGACATACATGAAGT
TCGTATTGGCTGGTCACTAACTACAAGTGGAATGTTACTTGGTGAAGAAGAATTTTCTTATGGGTATTCT
CTAAAAGGAATAAAAACATGCAACTGTGAGACTGAAGATTATGGAGAAAAGTTTGATGAAAATGATGTGA
TTACATGTTTTGCTAACTTTGAAAGTGATGAAGTAGAACTCTCGTATGCTAAGAATGGACAAGATCTTGG
CGTTGCCTTCAAAATCAGTAAGGAAGTTCTTGCTGGACGGCCACTGTTCCCGCATGTTCTCTGCCACAAC
TGTGCAGTTGAATTTAATTTTGGTCAGAAGGAAAAGCCATATTTTCCAATACCTGAAGAGTATACTTTCA
TCCAGAACGTCCCCTTAGAGGATCGAGTTAGAGGACCAAAGGGGCCTGAAGAGAAGAAAGATTGTGAAGT
TGTGATGATGATTGGCTTGCCAGGAGCTGGAAAAACTACCTGGGTTACTAAACATGCAGCAGAAAATCCA
GGGAAATATAACATTCTTGGCACAAATACTATTATGGATAAGATGATGGTGGCAGGTTTTAAGAAGCAAA
TGGCAGATACTGGAAAACTGAACACACTGTTGCAGAGAGCCCCCCAGTGTCTTGGGAAATTTATTGAGAT
TGCTGCCCGAAAGAAGCGAAATTTTATTCTGGATCAGACAAATGTGTCTGCTGCTGCCCAGAGGAGAAAA
ATGTGCCTGTTTGCAGGCTTCCAGCGAAAAGCTGTTGTAGTTTGCCCAAAAGATGAAGACTATAAGCAAA
GAACACAGAAGAAAGCAGAAGTAGAGGGGAAAGACCTACCAGAACATGCGGTCCTCAAAATGAAAGGAAA
CTTTACCCTCCCAGAGGTAGCTGAGTGCTTTGATGAAATAACCTATGTTGAACTTCAGAAGGAAGAAGCC
CAAAAACTCTTGGAGCAATATAAGGAAGAAAGCAAAAAGGCTCTTCCACCAGAAAAGAAACAGAACACTG
GCTCAAAGAAAAGCAATAAAAATAAGAGTGGCAAGAACCAGTTTAACAGAGGTGGTGGCCATAGAGGACG
TGGAGGATTCAATATGCGTGGTGGAAATTTCAGAGGAGGAGCCCCTGGGAATCGTGGCGGATATAATAGG
AGGGGCAACATGCCACAGAGAGGTGGTGGCGGTGGAGGAAGTGGTGGAATCGGCTATCCATACCCTCGTG
CCCCTGTTTTTCCTGGCCGTGGTAGTTACTCAAACAGAGGGAACTACAACAGAGGTGGAATGCCCAACAG
AGGGAACTACAACCAGAACTTCAGAGGACGAGGAAACAATCGTGGCTACAAAAATCAATCTCAGGGCTAC
AACCAGTGGCAGCAGGGTCAATTCTGGGGTCAGAAGCCATGGAGTCAGCATTATCACCAAGGATATTATT
GAATACCCAAATAAAACGAACTGATACATATTTCTCCAAAACCTTCACAAGAAGTCGACTGTTTTCTTTA
GTAGGCTAACTTTTTAAACATTCCACAAGAGGAAGTGCCTGCGGGTTCCTTTTTTTAGAAGCTTTGTGGGT
TGATTTTTTTTTCTTTTCTTTTTTGTACATTTTTAATTGCAGTTTAAAAGTGAATCGTAAGAGAACCTCAG
CATTGTGCACGATAAGAGAATGTGTCAGTATTTCAGGGTTCTACATTTTATCTGTAAAATGTGACTTTTT
TTTTTTTTTATCACAACAGAAGTAAAATGTTGCTTTGTACCTGGTGTCTTTTATTAAGAATTTACTCCCC
CCATTTCTCACAGAGAATAACAGTCGGGAGTCATTGTCACAATATAATAGAAATGTTAGCAACCAGATTC
ATGTAAGGACTAAGTGGTCCTCATGAATTGCATTAAGACTCTGTACTGCTCATATTACACTCCATCCTCT
CTGTAGTTTGCTGGGTAGTGGAGGGGGTAAGCTAAATCATAGTTTCTGACAATAACTGGGAAGGTTTTTT

CTTAAAATAACAATGGAATTGGTATAATTGGGATTGAAAACTAAAACTTGGAACTAAGATAGAGAAGATG
GAGTGTATGTAGAAGGGCTGTTAAAAATGTAAAACTTGGTTGCATTATTTGTGGAGGCTCAAACTTGTGA
AGGTTAATACCATAATTTTTCCATTTGTTCTGCATTTTGATTCTGAAAAGAAAGCTGGCTTTGCCCATTT
CTTATTAAAAAAACTTGTTGTAAATCCAGTTGTCTAATGGGATCTATATGAAGTTAGCCATGTCTGTATG
CCCTTCTCCCACAAAATACTGTATAACTAGTGTGCTTGTAGTAGTTAACTCCACCATCTTTGTAAGCTAA
TGAAATTGTGAGTCACCCATTTATATCTTAATTTTTAATCATGTCAGTTCTTGAATGGGTATCTCCTTAG
CCTGCTGATTTCTTTTTCTTTCTAAAGAAAGTGGGTGGAGAAATTAATTTAGACGTTTGTTTGCAATAAA
AAGAATTCATTTTA

>DEX0356_037.nt.1
GAGTATATATATATGTACTCCAAGTTCAATAAGGAAATTTTAAAAGGCACTCATCAGCATAGTGCTAGTT
GTAATTAAAGCAGGCATACTGGCATATAGGGCCTTTAATATTACAGGAATTAAGATGATCCTAACGTGTT
TTGTACTGGTCGAAGCTTATCCTTAGGTGACAGGTTCTAGAGTCTATAAAGATAATTTAGTAATTTTCAC
TTTTTGACTCATGGGATTTACAAGAATATTTAAAGGGCTTAGAAAGTAAAGACTTGCGAAGAAAAGCTTT
AGCAGGAGAGATTGTTCAATATAGAAGAGAAAAATATCCCAGTGAATTCCTGAAGGAAGGTATGGCATAT
TTAATAGCAGCCCTGATTTAAAATAGTGGGAATTAGGTGATCCTGTAGATCCTTCCAACTTTATGATTTT
TAGAGATTTTATTTTTGCCCAAGTCATCAGTAAGTATGAGGACATTTTTTTTGTGTTTTTCCAAGTGGGT
TGTTTAAAATTTTTTCCTTTTTTTATTCCTCTTCCCAAGGCAAATGAAAAAAAAGTTGACCAGCCTCCAG
AAGCTAAAAAGCCCAAAATAAAGGTGGTGAATGTTGAGCTGCCTATTGAAGCCAACTTGGTCTGGCAGTT
AGGGAAAGACCTTCTTAACATGTATATTGAGACAGAGGGTAAGATGATAATGCAAGATAAATTGGAAAAA
GAAAGGAATGATGCTAAAAATGCAGTTGAGGAATATGTGTATGAGTTCAGAGACAAGCTGTGTGGACCAT
ATGAAAAATTTATATGTGAGCAGGATCATCAAAATTTTTTGAGACTCCTCACAGAAACTGAAGACTGGCT
GTATGAAGAAGGAGAGGACCAAGCTAAACAAGCATATGTTGACAAGTTGGAAGAATTAATGAAAATTGGC
ACTCCAGTTAAAGTTCGGTTTCAGGAAGCTGAAGAACGGCCAAAAATGTTTGAAGAACTAGGACAGAGGC
TGCAGCATTATGCCAAGATAGCAGCTGACTTCAGAAATAAGGATGAGAAATACAACCATATTGATGAGTC
TGAAATGAAAAAAGTGGAGAAGTCTGTTAATGAAGTGATGGAATGGATGAATAATGTCATGAATGCTCAG
GCTAAAAAGAGTCTTGATCAGGATCCAGTTGTACGTGCTCAGGAAATTAAAACAAAAATCAAGGAATTGA
ACAACACATGTGAACCCGTTGTAACACAACCGAAACCAAAAATTGAATCACCCAAACTGGAAAGAACTCC
AAATGGCCCAAATATTGATAAAAAGGAAGAAGATTTAGAAGACAAAAACAATTTTGGTGCTGAACCTCCA
CATCAGAATGGTGAATGTTACCCTAATGAGAAAAATTCTGTTAATATGGACTTGGACTAGATAACCTTAA
ATTGGCCTATTCCTTCAATTAATAAAATATTTTTGCCATAGTATGTGACTCTACATAACATACTGAAACT
ATTTATATTTTCTTTTTTAAGGATATTTAGAAATTTTGTGTATTATATGGAAAAAGAAAAAAAGCTTAAG
TCTGTAGTCTTTATGATCCTAAAAGGGAAAATTGCCTTGGTAACTTTCAGATTCCTGTGGAATTGTGAAT
TCATACTAAGCTTTCTGTGCAGTCTCACCATTTGCATCACTGAGGATGAAACTGACTTTTGTCTTTTGGA
GAAAAAAAACTGTACTGCTTGTTCAAGAGGGCTGTGATTAAAATCTTTAAGCATTTGTTCCTGCCAAGGT
AGTTTTCTTGCATTTTGCTCTCCATTCAGCATGTGTGTGGGTGTGGATGTTTATAAACAAGACTAAGTCT
GACTTCATAAGGGCTTTCTAAAACCATTTCTGTCCAAGAGAAAATGACTTTTTGCTTTGATATTAAAAAT
TCAATGAGTAAAACAAAAGCTAGTCAAATGTGTTAGCAGCATGCAGAACAAAAACTTTAAACTTTCTCTC
TCACTATACAGTATATTGTCATGTGAAAGTGTGGAATGGAAGAAATGTCGATCCTGTTGTAACTGATTGT
GAACACTTTTATGAGCTTTAAAATAAAGTTCATCTTATGGTGTCATTTCTA

>DEX0356_038.nt.1
CCGCAAAAAGCAAAGCAGAAAGCCAAACAAGACCCAACCAACTTGGGTTATTCATTCGATTCAAAACTGT
TTTTGGTTAAATCAAACATGAAAGTACAAAACCGCAGGAACGCGCATCTTAGCTCATTTGATCGCTTCGC
CTTCCTTGGGAAATGCAGTTTCGTGTCACCTGTTGCAGAAGATATGTAGTTGATCATCTAGACATAATTG
CTGAAGATGAACTTTTGGACCAACTTCTAAGTCACACAGCATCATTATCAGATTTATTACAGAACTCGGT
GGTGGCCACTGCGCAGACCAGACTTCGCTCGTACTCGTGCGCCTCGCTTCGCTTTTCCTCCGCAACCATG
TCTGACAAACCCGATATGGCTGAGATCGAGAAATTCGATAAGTCGAAACTGAAGAAGACAGAGACGCAAG
AGAAAAATCCACTGCCTTCCAAAGAAACGATTGAACAGGAGAAGCAAGCAGGCGAATCGTAATGAGGCGT
GCGCCGCCAATATGCACTGTACATTCCACAAGCATTGCCTTCTTATTTTACTTCTTTTAGCTGTTTAACT
TTGTAAGATGCAAAGAGGTTGGATCAAGTTTAAATGACTGTGCTGCCCCTTTCACATCAAAGAACTACTG
ACAACGAAGGCCGCGCCTGCCTTTCCCATCTGTCTATCTATCTGGCTGGCAGGGAAGGAAAGAACTTGCA
TGTTGGTGAAGGAAGAAGTGGGGTGGAAGAAGTGGGGTGGGACGACAGTGAAATCTAGAAGTGCCATTTT
TTTTTTTGTTCAAATGATTTTAATTATTGGAATGCACAATTTTTTTAATATGCAAATAAAAAGTTTAAAA
ACTTA

>DEX0356_039.nt.1
TCTCCGCCTGCCGAGCTCCGACCCGGCGTTCCGAGTCCCCGGCGCCCGGAGGTGACGGTCGCTTCCCCTG
CCCTCCCTTAGGGTTAGTCCCGCCCAATCGGCTCCTTAATTTTTCGGAGGGGCTTCTACTTTGGCATTTA
TCACGGGCTGTCTGCCTCACTTACGAAGCAGGGTGCCCCTAGCACACTGCCGGGCCCCTGGGGGAACGCA

CTTAGGGAAGCGAAGCAAAGCTAACTGCATTTGGCAGAAACCAGCGCGCTTTCCCACGGCGCCACGGGAG
CGCACCGCTAGACAGAGAAGGCACTAGAAGAGGCGGTCGGTGCATAGCGTCACCTCCTAAACTCCGCCCA
TCGGAAGCGCCTTCTAAGCTCGTTCTTCCGCCAGCTTCCCTCCTCTTCCTTTCTCCGCCATCGTGGTGTG
TTCTTGACTCCGCTGCTCGCCATGTCTTCTCACAAGACTTTCAGGATTAAGCGATTCCTGGCCAAGAAAC
AAAAGCAAAATCGTCCCATTCCCCAGTGGATTCGGATGAAAACTGGAAATAAAATCAGCTCTTGAAGATG
ATGTGTCTCGGAGTTTGTTCCTTCAGATGTTCAGATATGTCCAGAGTTTCTTCCTTCCAGTGGGTTCGTG
GTCTCACTGACTTCAGGAATGAAGCTGCATACGCTCACGCTCCACAGAAAAGTTCTCCAAGTCCCCACTC
GACCCAGGAAGTCCTGCTGTCTTCACCTGTCACCAGCTACTGGGGAGGCTGAGGCAGGAGAATCGCTTGA
ACCCAGGAGGTGGAGGTTGCTGGGAGCCAAGCCAAGATCACGCCACTACACTCCAGCCTGGCAACAGAGT
GAGACTCCGTCTC

>DEX0356_040.nt.1
TACGGCTAGGTACGAGGCTGGGTGTTGCTCCTCTCCCAGTCGGCACCACAGCGGTGGCTGCCGGGCGTGG
TGTCGGTGGGTCGGTTGGTTTTTGTCTCACCGTTGGTGTCCGTGCCGTTCAGTTGCCCGCCATGGCTGAG
CTGGATCCGTTCGGCGCCCCTGCCGGCGCCCCTGGCGGTCCCGCGCTGGGGAACGGAGTGGCCGGCGCCG
GCGAAGAAGACCCGGCTGCGGCCTTCTTGGCGCAGCAAGAGAGCGAGATTGCGGGCATCGAGAACGACGA
GGCCTTCGCCATCCTGGACGGCGGCGCCCCCGGGCCCCAGCCGCACGGCGAGCCGCCGGGGGGTCCGGAT
GCTGTTGATGGAGTAATGAATGGTGAATACTACCAGGAAAGTAATGGTCCAACAGACAGTTATGCAGCTA
TTTCACAAGTGGATCGATTGCAGTCAGAGCCTGAAAGTATCCGTAAATGGAGAGAAGAACAAATGGAACG
CTTGGAAGCCCTTGATGCCAATTCTCGGAAGCAAGAAGCAGAGTGGAAAGAAAAGGCAATAAAGGAGCTA
GAAGAATGGTATGCAAGACAGGACGAGCAGCTACAGAAAACAAAAGCAAACAACAGGGCAGCAGAAGAAG
CCTTTGTAAATGACATTGACGAGTCGTCCCCAGGCACTGAGTGGGAACGGGTGGCCCGGCTGTGTGACTT
TAACCCCAAGTCTAGCAAGCAGGCCAAAGATGTCTCCCGCATGCGCTCAGTCCTCATCTCCCTCAAGCAG
GCCGCTGGTGCTGGAGGCCCCGTGGACCTCCTGGTTTTACCCCGGGGCCTCCGCTGGGTTTTGGGTTGGC
CCTTGTTGGCTTT

>DEX0356_040.nt.2
CTAGTTCTAGATCGCGGGCGGCCGCTCGCGATCTAGAACCGGACGCGTGGGGCGCTTGTCCTCCTCTCCC
AGTCGGCACCACAGCGGTGGCTGCCGGGCGTGGTGTCGGTGGGTCGGTTGGTTTTTGTCTCACCGTTGGT
GTCCGTGCCGTTCAGTTGCCCGCCATGGCTGAGCTGGATCCGTTCGGCGCCCCTGCCGGCGCCCCTGGCG
GTCCCGCGCTGGGGAACGGAGTGGCCGGCGCCGGCGAAGAAGACCCGGCTGCGGCCTTCTTGGCGCAGCA
AGAGAGCGAGATTGCGGGCATCGAGAACGACGAGGCCTTCGCCATCCTGGACGGCGGCGCCCCCGGGCCC
CAGCCGCACGGCGAGCCGCCGGGGGGTCCGGATGCAGGAAAGTAATGGTCCAACAGACAGTTATGCAGCT
ATTTCACAAGTGGATCGATTGCAGTCAGAGCCTGAAAGTATCCGTAAATGGAGAGAAGAACAAATGGAAC
GCTTGGAAGCCCTTGATGCCAATTCTCGGAAGCAAGAAGCAGAGTGGAAAGAAAAGGCAATAAAGGAGCT
AGAAGAATGGTATGCAAGACAGGACGAGCAGCTACAGAAAACAAAAGCAAACAACAGGGCAGCAGAAGAA
GCCTTTGTAAATGACATTGACGAGTCGTCCCCAGGCACTGAGTGGGAACGGGTGGCCCGGCTGTGTGACT
TTAACCCCAAGTCTAGCAAGCAGGCCAAAGATGTCTCCCGCATGCGCTCAGTCCTCATCTCCCTCAAGCA
GGCCCCGCTGGTGCACTGAAGAGCCACCCTGTGGAAACACTACATCTGCAATATCTTAATCCTACTCAGT
GAAGCTCTTCACAGTCATTGGATTAATTATGTTGAGTTCTTTTGGACCAAACCTTTTTGTCTTTAGAGTT
GTTCATTGTTTGTGATTGCATGTTTCCTTCCTTCAACTGTGTTCTCCCTGGCATTCAGAGAGGAGGGAGA
GGAGGAAGAGGAAGGGGAGGGAAGCTTCCCAAGAGTAGCCTCAACCTGTGCTTCTGTGCATTATTCTGAG
AATAAATTTCTGTTTCAAACTGTAAAAAAAAAAAAAAACACAAAACACAACACACAAAAAACACAAAAACA
CAAAAAAAAAAACAAACAACACACAAAACAAACACACAACACACACCACACACAAACAAAAAACAAACAC
AAAGAGGGGGGGGGCGCATATATATATAATAACCACCCCGCGGGAGGGGCACAATATTATCATCACACAC
AACACCACAATATTATGAGAAAAAAAGGCGCCCCCAATTAAGAGAGGGGTGTATATAATAAAAAAAGAGA
AGAGGGCCCACTTTTAATATACACAACCGGCTAAGAGCGGAGAAGAAACAAACACACGCTTGTCCGCCAG
GAGAACTCTTTTGGTGGAAAGACACCCACCCCATCTCTCCTGCGGGGGAGGCAAACAAATCAATGTGGAC
CAACACCACACCCCTCCAGAAGAGGAGATTATAACCCCCCCCCTCCACGGAGGGATCATCAACAACACCA
ACACAGTGAGAGAGATCATAAGGGTGGACACAAAAACCGCGACCATACAATCAGCTGATGAGGAGCTGAA
AAAAAATCATTGTCATCACACCGGGCGTTCAATTCTAAAAAAAAATTATATAGAGA

>DEX0356_040.nt.3
CTAGTTCTAGATCGCGGGCGGCCGCTCGCGATCTAGAACCGGACGCGTGGGGCGCTTGTCCTCCTCTCCC
AGTCGGCACCACAGCGGTGGCTGCCGGGCGTGGTGTCGGTGGGTCGGTTGGTTTTTGTCTCACCGTTGGT
GTCCGTGCCGTTCAGTTGCCCGCCATGGCTGAGCTGGATCCGTTCGGCGCCCCTGCCGGCGCCCCTGGCG
GTCCCGCGCTGGGGAACGGAGTGGCCGGCGCCGGCGAAGAAGACCCGGCTGCGGCCTTCTTGGCGCAGCA
AGAGAGCGAGATTGCGGGCATCGAGAACGACGAGGCCTTCGCCATCCTGGACGGCGGCGCCCCCGGGCCC
CAGCCGCACGGCGAGCCGCCGGGGGGTCCGGATGCCAATTCTCGGAAGCAAGAAGCAGAGTGGAAAGAAA
AGGCAATAAAGGAGCTAGAAGAATGGTATGCAAGACAGGACGAGCAGCTACAGAAAACAAAAGCAAACAA

CAGGGCAGCAGAAGAAGCCTTTGTAAATGACATTGACGAGTCGTCCCCAGGCACTGAGTGGGAACGGGTG
GCCCGGCTGTGTGACTTTAACCCCAAGTCTAGCAAGCAGGCCAAAGATGTCTCCCGCATGCGCTCAGTCC
TCATCTCCCTCAAGCAGGCCCCGCTGGTGCACTGAAGAGCCACCCTGTGGAAACACTACATCTGCAATAT
CTTAATCCTACTCAGTGAAGCTCTTCACAGTCATTGGATTAATTATGTTGAGTTCTTTTGGACCAAACCT
TTTTGTCTTTAGAGTTGTTCATTGTTTGTGATTGCATGTTTCCTTCCTTCAACTGTGTTCTCCCTGGCAT
TCAGAGAGGAGGGAGAGGAGGAAGAGGAAGGGGAGGGAAGCTTCCCAAGAGTAGCCTCAACCTGTGCTTC
TGTGCATTATTCTGAGAATAAATTTCTGTTTCAAACTGTAAAAAAAAAAAAAAAACACAAAACACAACACA
CAAAAACACAAAAACACAAAAAAAAAAACAAACAACACACAAAACAAACACACAACACACACCACACACA
AACAAAAAACAAACACAAAGAGGGGGGGGGCGCATATATATATAATAACCACCCCGCGGGAGGGGCACAA
TATTATCATCACACACAACACCACAATATTATGAGAAAAAAAGGCGCCCCCAATTAAGAGAGGGGTGTAT
ATAATAAAAAAAAGAGAAGAGGGCCCACTTTTAATATACACAACCGGCTAAGAGCGGAGAAGAAACAAACA
CACGCTTGTCCGCCAGGAGAACTCTTTTGGTGGAAAGACACCCACCCCATCTCTCCTGCGGGGGAGGCAA
ACAAATCAATGTGGACCAACACCACACCCCTCCAGAAGAGGAGATTATAACCCCCCCCCTCCACGGAGGG
ATCATCAACAACACCAACACAGTGAGAGAGATCATAAGGGTGGACACAAAAACCGCGACCATACAATCAG
CTGATGAGGAGCTGAAAAAAAATCATTGTCATCACACCGGGCGTTCAATTCTAAAAAAAAATTATATAGA
GA

>DEX0356_040.nt.4
GCAGACGGGGTGGGTAGGACGCTCCCTATCAGGCAGCACTTCCGCCTCCCGGGGCCCGCGCAGCTCACCT
CCCTCACCTCCCGCCCTACCCCAGTCACGAGTTGTTTTAGGGGGACCGCCCCTCCACTTGCTGATTGGGT
AGCTCCTGAACCATTGTTGTCCTCTGATTGGTTGTTCCCTTTTCGGCTCTGCAACACCGCCTAGACCGAC
CGGATACACGGGTAGGGCTTCCGCTTTACCCGTCTCCCTCCTGGCGCTTGTCCTCCTCTCCCAGTCGGCA
CCACAGCGGTGGCTGCCGGGCGTGGTGTCGGTGGGTCGGTTGGTTTTTGTCTCACCGTTGGTGTCCGTGC
CGTTCAGTTGCCCGCCATGGCTGAGCTGGATCCGTTCGGCGCCCCTGCCGGCGCCCCTGGCGGTCCCGCG
CTGGGGAACGGAGTGGCCGGCGCCGGCGAAGAAGACCCGGCTGCGGCCTTCTTGGCGCAGCAAGAGAGCG
AGATTGCGGGCATCGAGAACGACGAGGCCTTCGCCATCCTGGACGGCGGCGCCCCCGGGCCCCAGCCGCA
CGGCGAGCCGCCGGGGGGTCCGGATGCTGTTGATGGAGTAATGAATGGTGAATACTACCAGGAAAGTAAT
GGTCCAACAGACAGTTATGCAGCTATTTCACAAGTGGATCGATTGCAGTCAGAGCCTGAAAGTATCCGTA
AATGGAGAGAAGAACAAATGGAACGCTTGGAAGCCCTTGATGCCAATTCTCGGAAGCAAGAAGCAGAGTG
GAAAGAAAAGGCAATAAAGGAGCTAGAAGAATGGTATGCAAGACAGGACGAGCAGCTACAGAAAACAAAA
GCAAACAACAGGGTGGCAGATGAAGCTTTCTACAAACAACCCTTCGCTGACGTGATTGGTTATGTGCAGG
GCAGCAGAAGAAGCCTTTGTAAATGACATTGACGAGTCGTCCCCAGGCACTGAGTGGGAACGGGTGGCCC
GGCTGTGTGACTTTAACCCCAAGTCTAGCAAGCAGGCCAAAGATGTCTCCCGCATGCGCTCAGTCCTCAT
CTCCCTCAAGCAGGCCCCGCTGGTGCACTGAAGAGCCACCCTGTGGAAACACTACATCTGCAATATCTTA
ATCCTACTCAGTGAAGCTCTTCACAGTCATTGGATTAATTATGTTGAGTTCTTTTGGACCAAACCTTTTT
GTCTTTAGAGTTGTTCATTGTTTGTGATTGCATGTTTCCTTCCTTCAACTGTGTTCTCCCTGGCATTCAG
AGAGGAGGGAGAGGAGGAAGAGGAAGGGGAGGGAAGCTTCCCAAGAGTAGCCTCAACCTGTGCTTCTGTG
CATTATTCTGAGAATAAATTTCTGTTTCAAACTGTAAAAAAAAAAAAAAAACACAAAACACAACACACAAA
AACACAAAAACACAAAAAAAAAAACAAACAACACACAAAACAAACACACAACACACACCACACACAAACA
AAAAACAAACACAAAGAGGGGGGGGGCGCATATATATATAATAACCACCCCGCGGGAGGGGCACAATATT
ATCATCACACACAACACCACAATATTATGAGAAAAAAAGGCGCCCCCAATTAAGAGAGGGGTGTATATAA
TAAAAAAAGAGAAGAGGGCCCACTTTTAATATACACAACCGGCTAAGAGCGGAGAAGAAACAAACACACG
CTTGTCCGCCAGGAGAACTCTTTTGGTGGAAAGACACCCACCCCATCTCTCCTGCGGGGGAGGCAAACAA
ATCAATGTGGACCAACACCACACCCCTCCAGAAGAGGAGATTATAACCCCCCCCCTCCACGGAGGGATCA
TCAACAACACCAACACAGTGAGAGAGATCATAAGGGTGGACACAAAAACCGCGACCATACAATCAGCTGA
TGAGGAGCTGAAAAAAAATCATTGTCATCACACCGGGCGTTCAATTCTAAAAAAAAATTATATAGAGA

>DEX0356_040.nt.5
GCAGACGGGGTGGGTAGGACGCTCCCTATCAGGCAGCACTTCCGCCTCCCGGGGCCCGCGCAGCTCACCT
CCCTCACCTCCCGCCCTACCCCAGTCACGAGTTGTTTTAGGGGGACCGCCCCTCCACTTGCTGATTGGGT
AGCTCCTGAACCATTGTTGTCCTCTGATTGGTTGTTCCCTTTTCGGCTCTGCAACACCGCCTAGACCGAC
CGGATACACGGGTAGGGCTTCCGCTTTACCCGTCTCCCTCCTGGCGCTTGTCCTCCTCTCCCAGTCGGCA
CCACAGCGGTGGCTGCCGGGCGTGGTGTCGGTGGGTCGGTTGGTTTTTGTCTCACCGTTGGTGTCCGTGC
CGTTCAGTTGCCCGCCATGGCTGAGCTGGATCCGTTCGGCGCCCCTGCCGGCGCCCCTGGCGGTCCCGCG
CTGGGGAACGGAGTGGCCGGCGCCGGCGAAGAAGACCCGGCTGCGGCCTTCTTGGCGCAGCAAGAGAGCG
AGATTGCGGGCATCGAGAACGACGAGGCCTTCGCCATCCTGGACGGCGGCGCCCCCGGGCCCCAGCCGCA
CGGCGAGCCGCCGGGGGGTCCGGATGCTGTTGATGGAGTAATGAATGGTGAATACTACCAGGAAAGTAAT
GGTCCAACAGACAGTTATGCAGCTATTTCACAAGTGGATCGATTGCAGTCAGAGCCTGAAAGTATCCGTA
AATGGAGAGAAGAACAAATGGAACGCTTGGAAGCCCTTGATGCCAATTCTCGGAAGCAAGAAGCAGAGTG
GAAAGAAAAGGCAATAAAGGAGCTAGAAGAATGGTATGCAAGACAGGACGAGCAGCTACAGAAAACAAAA

GCAAACAACAGGGCAGCAGAAGAAGCCTTTGTAAATGACATTGACGAGTCGTCCCCAGGCACTGAGTGGG
AACGGGTGGCCCGGCTGTGTGACTTTAACCCCAAGTCTAGCAAGCAGGCCAAAGATGTCTCCCGCATGCG
CTCAGTCCTCATCTCCCTCAAGCAGGCCCGKKGGTGATGARRAGCACCTGTGGAAMACTAATTTGATATT
TATCTATCAGGGAAGTTTAAGCATGATATATGTTGAGGTTTTGGAGAACCTTTGTTCGAGTGTCTTGTTG
GTGGGGTTCTCTAAGGGTTCCGGTAAGAGGGGGAGAAAGGGGGCAAAACGTTGGTTTAAATTTGGGAGAA
AAAAAAAGGGTTTGGGGTGTTTTGTGGGGTTGAAATTTATCNA

>DEX0356_040.nt.6
GCAGACGGGGTGGGTAGGACGCTCCCTATCAGGCAGCACTTCCGCCTCCCGGGGCCCGCGCAGCTCACCT
CCCTCACCTCCCGCCCTACCCCAGTCACGAGTTGTTTTAGGGGGACCGCCCCTCCACTTGCTGATTGGGT
AGCTCCTGAACCATTGTTGTCCTCTGATTGGTTGTTCCCTTTTCGGCTCTGCAACACCGCCTAGACCGAC
CGGATACACGGGTAGGGCTTCCGCTTTACCCGTCTCCCTCCTGGCGCTTGTCCTCCTCTCCCAGTCGGCA
CCACAGCGGTGGCTGCCGGGCGTGGTGTCGGTGGGTCGGTTGGTTTTTGTCTCACCGTTGGTGTCCGTGC
CGTTCAGTTGCCCGCCATGGCTGAGCTGGATCCGTTCGGCGCCCCTGCCGGCGCCCCTGGCGGTCCCGCG
CTGGGGAACGGAGTGGCCGGCGCCGGCGAAGAAGACCCGGCTGCGGCCTTCTTGGCGCAGCAAGAGAGCG
AGATTGCGGGCATCGAGAACGACGAGGCCTTCGCCATCCTGGACGGCGGCGCCCCCGGGCCCCAGCCGCA
CGGCGAGCCGCCGGGGGGTCCGGATGCTGTTGATGGAGTAATGAATGGTGAATACTACCAGGAAAGTAAT
GGTCCAACAGACAGTTATGCAGCTATTTCACAAGTGGATCGATTGCAGTCAGAGCCTGAAAGTATCCGTA
AATGGAGAGAAGAACAAATGGAACGCTTGGAAGCCCTTGATGCCAATTCTCGGAAGCAAGAAGCAGAGTG
GAAAGAAAAGGCAATAAAGGAGCTAGAAGAATGGTATGCAAGACAGGACGAGCAGCTACAGAAAACAAAA
GCAAACAACAGGGCAGCAGAAGAAGCCTTTGTAAATGACATTGACGAGTCGTCCCCAGGCACTGAGTGGG
AACGGGTGGCCCGGCTGTGTGACTTTAACCCCAAGTCTAGCAAGCGGGCAAGATGTTTCCCGATGGCTCA
GTCTCATTCCTCAAGGAGGCCCGTGTGCCTGAGAGCCCTGTGGAACCTACATTGATATTTATCTATCAGT
GAGTCTCAAGTCATGTTATTATGTTGAGTTCTTTGGCCAACCTTTTGTTTTAGAGTGTCTGGTTGGGATG
CTGGTCCTCTCAAGGTTTCCGGCTCAGAGGGGGGGGAGGGAGGGGGGCCCAGAVCCCCGGTGGGTTTGA
AATTGTGGGAAAAAACGTTTKGGRRGAKGGGTGTCCTTTTGCTTAGCCTTTTATGGGCCGGATCTATAAC
CN

>DEX0356_041.nt.1
AGCTCACCCTGGTCAGCCTAGCACTCTGACCTAGCAGTCAACATGAAGGCTCTCATTGTTCTGGGGCTTG
TCCTCCTTTCTGTTACGGTCCAGGGCAAGGTCTTTGAAAGGTGTGAGTTGGCCAGAACTCTGAAAAGATT
GGGAATGGATGGCTACAGGGGAATCAGCCTAGCAAACTGGATGTGTTTGGCCAAATGGGAGAGTGGTTAC
AACACACGAGCTACAAACTACAATGCTGGAGACAGAAGCACTGATTATGGGATATTTCAGATCAATAGCC
GCTACTGGTGTAATGATGGCAAAACCCCAGGAGCAGTTAATGCCTGTCATTTATCCTGCAGTGCTTTGCT
GCAAGATAACATCGCTGATGCTGTAGCTTGTGCAAAGAGGGTTGTCCGTGATCCACAAGGCATTAGAGCA
TGGGTGGCATGGAGAAATCGTTGTCTTTCCTTCTTCAGCTCATTTTGTCTCTCTCACATTAAGGGAGTAG
GAATTAAGTGAAAGGTCACACTACCATTATTTCCCCTTCAAACAAATAATATTTTTACAGAAGCAGGAGC
AAAATATGGCCTTTCTTCTAAGAGATATAATGTTCACTAATGTGGTTATTTTATATTAAGCCTACAACAT
TTTTCAGTTTGCAAATAGAACTAATACTGGTGAAAATTTACCTAAAACCTTGGTTATCAAATACATCTCC
AGTACATTCCGTTCTTTTTTTTTTTGAGACAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGTGGCGCAA
TCTCGGCTCACTGCAACCTCCACCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGG
GATTACGGGCGCCCGCCACCACGCCCGGCTAATTTTTTGTATTTTTAGTAGAGACAGGGTTTCACCGTGT
TAGCCAGGATGGTCTCGATCTCCTGACCTTGTGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTAC
AGGCGTGAGCCACTGCGCCCGGCCACATTCAGTTCTTATCAAAGAAATAACCCAGACTTAATCTTGAATG
ATACGATTATGCCCAATATTAAGTAAAAAATATAAGAAAAGGTTATCTTAAATAGATCTTAGGCAAAATA
CCAGCTGATGAAGGCATCTGATGCCTTCATCTGTTCAGTCATCTCCAAAAACAGTAAAAATAACCACTTT
TTGTTGGGCAATATGAAATTTTTAAAGGAGTAGAATACCAAATGATAGAAACAGACTGCCTGAATTGAGA
ATTTTGATTTCTTAAAGTGTGTTTCTTTCTAAATTGCTGTTCCTTAATTTGATTAATTTAATTCATGTAT
TATGATTAAATCTGAGGCAGATGAGCTTACAAGTATTGAAATAATTACTAATTAATCACAAATGTGAAGT
TATGCATGATGTAAAAAATACAAACATTCTAATTAAAGGCTTTGCAACACA

>DEX0356_042.nt.1
GTCTGAGGGAGGGGCCTGGACTCCGGGCTCTGAGCCTACTGGGGCACAGGCCTGTGAGTTGAGGCCCCTT
CTCATCATCCCCCTCCTCAGGTGGTCTACATAGCCTGCTCCTTCACCACGGTCTGGTTGATTTATAGCAA
GTTCAAAGCTACTTACGATGGGAACCATGACACGTTCAGAGTGGAGTTCCTGGTCGTTCCCACAGCCATT
CTGGCGTTCCTGGTCAATCATGACTTCACCCCTCTGGAGATCCTCTGGACCTTCTCCATCTACCTGGAGT
CAGTGGCCATCTTGCCGCAGCTGTTCATGGTGAGCAAGACCGGCGAGGCGGAGACCATCACCAGCCACTA
CTTGTTTGCGCTAGGCGTTTACCGCACGCTCTATCTCTTCAACTGGATCTGGCGCTACCATTTCGAGGGC
TTCTTCGACCTCATCGCCATTGTGGCAGGCCTGGTCCAGACAGTCCTCTACTGCGATTTCTTCTACCTCT
ATATCACCAAAGTCCTAAAGGGGAAGAAGTTGAGTTTGCCGGCATAGCCCCGGTCCTCTCCATCTCTCTC

CTCGGCAGCAGCGGGAGGCAGAGGAAGGCGGCAGAAGATGAAGAGCTTTCCCATCCAGGGGTGACTTTTT
TAAGAACCCACCTCTTGTGCTCCCCATCCCGCCTCCTGCCGGGTTTCAGGGGGACAGTGGAGGATCCAGG
TCTTGGGGAGCTCAGGACTTGGGCTGTTTGTAGTTTTTTGCCTTTTAGACAAGAAAAAAAAATCTTTCCA
CTCTTTAGTTTTTGATTCTGATGACTCGTTTTTCTTCTACTCTGTGGGCCCAATTTTTATAAAGTGTTTT
TGAGTGTCCTATGGGCCGGGGCAGGGTCCAAGATCTTTTCCCTTCCCCAGGCCCCTCGGCTCCCTCCCAG
ATCCCACCCCCAGCCCCACTGGTTGCCAAACACTAAATCTGCCGACACCCATCTGCCCCACCTCCTGCCA
TGGCCATGAACCGCGACCCCCACTAAATTTCTAGATTGGGGATAGGGAGAAAGGGAGGCCCAGGAAGGTC
TCCCCTGATTTTTTTTCATAGTAATTTTTTTCCCCAGAGTTTGAATTTTTTGGTCTTCTCCTGGTTTTTT
GGCAAATTAGGGGGGCCCGGGGCTCAAGTGCGGGAAGGGGGCTGGCCCGAGGATCCCATGGCTCTCACAC
CATGTTTTTGTACAGAACTGATGGTTGAATCTTTGTTCTCTTGAAATAAACAGAAGAAAATGAAACCTT

>DEX0356_043.nt.1
AGCTGACCTAGAGGATTYCCAGGCAAAGCYCTACGGCGCGTACAAGCAACGCGCAGCCTGGATCATCCCA
TTCTTCTGTTACCAGATCTTTGACTTTGCCCTGAACATGTTGGTTGCAATCACTGTGCTTATTTATCCAA
ACTCCATTCAGGAATACATACGGCAACTGCCTCCTAATTTTCCCTACAGAGATGATGTCATGTCAGTGAA
TCCTACCTGTTTGGTCCTTATTATTCTTCTGTTTATTAGCATTATCTTGACTTTTAAGGGTTACTTGATT
AGCTGTGTTTGGAACTGCTACCGATACATCAATGGTAGGAACTCCTCTGATGTCCTGGTTTATGTTACCA
GCAATGACACTACGGTGCTGCTACCCCCGTATGATGATGCCACTGTGAATGGTGCTGCCAAGGAGCCACC
GCCACCTTACGTGTCTGCCTAAGCCTTCAAGTGGGCGGAGCTGAGGGCAGCAGCTTGACTTTGCAGACAT
CTGAGCAATAGTTCTGTTATTTCACTTTTGCCATGAGCCTCTCTGAGCTTGTTTGTTGCTGAAATGCTAC
TTTTTAAAATTTAGATGTTAGATTGAAAACTGTAGTTTTCAACATATGCTTTGCTAGAACACTGTGATAG
ATTAACTGTAGAATTCTTCCTGTACGATTGGGGATATAACGGGCTTCACTAACCTTCCCTAGGCATTGAA
ACTTCCCCCAAATCTGATGGACCTAGAAGTCTGCTTTTGTACCTGCTGGGCCCCAAAGTTGGGCATTTTT
CTCTCTGTTCCCTCTCTTTTGAAAATGTAAAATAAAACCAAAAATAGACAACTTTTTCTTCAGCCATTCC
AGCATAGAGAACAAAACCTTATGGAAACAGGAATGTCAATTGTGTAATCATTGTTCTAATTAGGTAAATA
GAAGTCCTTATGTATGTGTTACAAGAATTTCCCCCACAACATCCTTTATGACTGAAGTTCAATGACAGTT
TGTGTTTGGTGGTAAAGGATTTTCTCCATGGCCTGAATTAAGACCATTAGAAAGCACCAGGCCGTGGGAG
CAGTGACCATCTGCTGACTGTTCTTGTGGATCTTGTGTCCAGGGACATGGGGTGACATGCCTCGTATGTG
TTAGAGGGTGGAATGGATGTGTTTGGCGCTGCATGGGATCTGGTGCCCCTCTTCTCCTGGATTCACATCC
CCACCCAGGGCCCGCTTTTACTAAGTGTTCTGCCCTAGATTGGTTCAAGGAGGTCATCCAACTGACTTTA
TCAAGTGGAATTGGGATATATTTGATATACTTCTGCCTAACAACATGGAAAAGGGTTTTCTTTTCCCTGC
AAGCTACATCCTACTGCTTTGAACTTCCAAGTATGTCTAGTCACCTTTTAAAATGTAAACATTTTCAGAA
AAATGAGGATTGCCTTCCTTGTATGCGCTTTTTACCTTGACTACCTGAATTGCAAGGGATTTTTATATAT
TCATATGTTACAAAGTCAGCAACTCTCCTGTTGGTTCATTATTGAATGTGCTGTAAATTAAGTCGTTTGC
AATTAAAACAAGGTTTGCCCCA

>DEX0356_043.nt.2
AGTCCGCCCCGCCCCCTCCCCGTCCCCGCCGCTGCAGCCGTCGCCTTCGGAGCGAAGGGTACCGACCCGG
CAGAAGCTCGGAGCTCTCGGGGTATCGAGGAGGCAGGCCCGCGGGCGCACGGGCGAGCGGGCCGGGAGCC
GGAGCGGCGGAGGAGCCGGCAGCAGCGGCGCGGCGGGCTCCAGGCGAGGCGGTCGACGCTCCTGAAAACT
TGCGCGCGCGCTCGCGCCACTGCGCCCGGAGCGATGAAGATGGTCGCGCCCTGGACGCGGTTCTACTCCA
ACAGCTGCTGCTTGTGCTGCCATGTCCGCACCGGCACCATCCTGCTCGGCGTCTGGTATCTGATCATCAA
TGCTGTGGTACTGTTGATTTTATTGAGTGCCCTGGCTGATCCGGATCAGTATAACTTTTCAAGTTCTGAA
CTGGGAGGTGACTTTGAGTTCATGGATGATGCCAACATGTGCATTGCCATTGCGATTTCTCTTCTCATGA
TCCTGATATGTGCTATGGCTACTTACGGAGCGTACAAGCAACGCGCAGCCTGGATCATCCCATTCTTCTG
TTACCAGATCTTTGACTTTGCCCTGAACATGTTGGTTGCAATCACTGTGCTTATTTATCCAAACTCCATT
CAGGAATACATACGGCAACTGCCTCCTAATTTTCCCTACAGAGATGATGTCATGTCAGTGAATCCTACCT
GTTTGGTCCTTATTATTCTTCTGTTTATTAGCATTATCTTGACTTTTAAGGGTTACTTGATTAGCTGTGT
TTGGAACTGCTACCGATACATCAATGGTAGGAACTCCTCTGATGTCCTGGTTTATGTTACCAGCAATGAC
ACTACGGGTTGTACCCGTTGATGATGCCTGTAATGTGTGCAGGAGCCGGCCTTGTGTTGGTAGCTAAGGG
GGAAGGGAGAGTGATTG

>DEX0356_044.nt.1
GAATTCGTGGGTCGTGCTGCGTCGACAACGGTAGTGACGCGTATTGCCTGGAGGATGGCGGACGCCGGCA
TTCGCCGCGTGGTTCCCAGCGACCTGTATCCCCTCGTGCTCGGCTTCCTGCGCGATAACCAACTCTCAGA
GGTGGCCAATAAGTTCGCCAAAGCGACAGGAGCTACACAGCAGGATGCCAATGCCTCTTCCCTCTTAGAC
ATCTATAGCTTCTGGCTCAACAGGTCTGCCAAGGTCCCAGAGCGAAAGTTACAGGCAAATGGACCAGTGG
CTAAGAAAGCTAAGAAGAAGGCCTCATCCAGTGACAGTGAGGACAGCAGCGAGGAGGAGGAGGAAGTTCA
AGGGCCTCCAGCAAAGAAGGCTGCTGTACCTGCCAAGCGAGTCGGTCTGCCTCCTGGGAAGGCTGCAGCC
AAAGCATCAGAGAGTAGCAGCAGTGAAGAGTCCAGTGATGATGATGATGAGGAGGACCAAAAGAAACAGC

CTGTCCAGAAGGGAGTTAAGCCCCAAGCCAAGGCAGCCAAAGCTCCTCCTAAGAAGGCCAAGAGCTCTGA
TTCTGATTCTGACTCAAGCTCCGAGGATGAGCCACCAAAGAACCAGAAGCCAAAGATAACACCTGTGACA
GTTAAAGCTCAGACTAAAGCCCCTCCCAAACCAGCTCGAGCAGCACCTAAAATAGCCAATGGTAAAGCAG
CCAGTAGCAGCAGTAGCAGCAGCAGCAGCAGTAGCAGTGATGACTCAGAGGAGGAGAAGGCAGCAGCCAC
CCCCAAGAAGGTCTGGACCATAACTTCTGTCAGGGCAGAGGTGACCAGGGTGTGATGTGTGTGTCTTC
CTTCCCTGGCAGGATTGGTTGGGTCAGCATTTTCCTGTGGTAATCAATTTTCTTTCTAATGCAGACTGTA
CCTAAAAAGCAAGTTGTGGCCAAGGCCCCAGTGAAAGCAGCTACCACCCCTACCCGGAAGAGTTCTAGCA
GTGAGGATTCCTCCAGTGACGAGGAAGAGGAGCAAAAAAAACCCATGAAAAATAAACCAGGTCCCTACAG
TTCAGTCCCCCCGCCTTCTGCTCCCCCACCAAAGAAGTCTCTGGGAACCCAGCCTCCCAAGAAGGCTGTG
GAGAAGCAGCAGCCTGTGGAAAGCAGTGAAGACAGCAGTGATGAGTCTGATTCAAGTTCTGAAGAAGAGA
AGAAACCCCCAACTAAGGCAGTAGTCTCTAAAGCAACCACTAAACCACCTCCAGCAAAGAAAGCAGCAGA
GAGCTCTTCAGACAGCTCAGACTCTGACAGCTCTGAGGATGATGAAGCTCCTTCTAAGCCAGCTGGTACC
ACCAAGAATTCTTCAAATAAGCCAGCTGTCACCACCAAGTCACCTGCAGTGAAGCCAGCTGCAGCCCCCA
AGCAACCTGTGGGCGGTGGCCAGAAGCTTCTGACGAGAAAGGCTGACAGCAGCTCCAGTGAGGAAGAGAG
CAGCTCCAGTGAGGAGGAGAAGACAAAGAAGATGGTGGCCACCACTAAGCCCAAGGCGACTGCCAAAGCA
GCTCTATCTCTGCCTGCCAAGCAGGCTCCTCAGGGTAGTAGGGACAGCAGCTCTGATTCAGACAGCTCCA
GCAGTGAGGAGGAGGAAGAGAAGACATCTAAGTCTGCAGTTAAGAAGAAGCCACAGAAGGTAGCAGGAGG
TGCAGCCCCTTCCAAGCCAGCCTCTGCAAAGAAAGGAAAGGCTGAGAGCAGCAACAGTTCTTCTTCTGAT
GACTCCAGTGAGGAAGAGGAAGAGAAGCTCAAGGGCAAGGGCTCTCCAAGACCACAAGCCCCCAAGGCCA
ATGGCACCTCTGCACTGACTGCCCAGAATGGAAAAGCAGCTAAGAACAGTGAGGAGGAGGAAGAAGAAAA
GAAAAAGGCGGCAGTGGTAGTTTCCAAATCAGGTTCATTAAAGAAGCGGAAGCAGAATGAGGCTGCCAAG
GAGGCAGAGACTCCTCAGGCCAAGAAGATAAAGCTTCAGACCCCTAACACATTTCCAAAAAGGAAGAAAG
GAGAAAAAAAGGGCATCATCCCCATTCCGAAGGGTCAGGGAGGAGGAAATTGAGGTGGATTCACGAGTTGC
GGACAACTCCTTTGATGCCAAGCGAGGTGCAGCCGGAGACTGGGGAGAGCGAGCCAATCAGGTTTTGAAG
TTCACCAAAGGCAAGTCCTTTCGGCATGAGAAAACCAAGAAGAAGCGGGGCAGCTACCGGGGAGGCTCAA
TCTCTGTCCAGGTCAATTCTATTAAGTTTGACAGCGAGTGACCTGAGGCCATCTTCGGTGAAGCAAGGGT
GATGATCGGAGACTACTTACTTTCTCCAGTGGACCTGGGAACCCTCAGGTCTCTAGGTGAGGGTCTTGAT
GAGGACAGAAGTTTAGAGTAGGTCCTAAGACTTTACAGTGTAACATCCTCTCTGGTCCTTTTCTGTGTTC
CTAGTTTTGTACAGACTTGTTTTTGAGTGTTGAGTAGCAGGGACAAAATAAGGGAATGTTATTTTTTAAG
AAAATTCATTTTCATTGTTGTCTCCTTCCTTTTCTGTGAAAGTCCTCATACTGAGAAATTTGTATATTTT
ATATTAAATCACTTACTATTGATTTTTGTTGTGATTTTCAAAGGTGGATTCCCACAGATAAAATCTTGGC
TATTGCCCAAAACATAGTAAAGGGTCACGTGTGACTTTTTATAATAGGAAGAAAATTCTGCCTTTGTGAG
TGCACATGTCCACATTTCATCCCTCCTTCCCTCAAAACCCTAGAGAGGGGCATTAAAGAATTGTTGATGT
ATATGCAATGTCTGTTAAGCATGCACTATGTATTTCATCCTCATTTATTGGGTCTGGGACTGAAGTTTTT
AGCCAGCATGGACCTAACCTACTTTTTGGGATAAAATTCTCTGTTTTGTTACAGGCAAAATTCTGGTATG
GCGTGAATGCCATGGGTCATTCTGAATATATTTTTTTCTGTAATTTTATCATTACACGATGTTTGCAATA
CGTGCTTTGTTTTTTAATTTGAAAGCAAACTTTTCTACTGTTGAAAGACATTTTTTGACAACTTGACCCT
TCCTAGTATTGAGTTCTAAGTTGAGGACTGCATCTTCTCGTTTTTTACAGTATAGAGAACAAAATGACAT
GAGTTTGAAAAATACATATCACTTGGTATTGCTGTCTTGGTTGCAGTGGTGATACAGAATTGGTTTCATT
AATTCCTACATGGTTGAGAATCACTGATCAAGAAAGTGGGGGGAAAAAAAACAAACGTTAAAACCTCAAT
CCTCAGTAGGAAGGTAGATTACATTAGGTGAAATTATAGGTAATCTATGTATGTGCTAATGGGGTTGGAA
AGAACCTTACAGAGCATATTACCTGATAAACTGGAGTGGGTTTGGGAGAACAAACTAATAGGATTATTGT
GTCTCCTAGTTGGTACCTGGGAGCAATTGACATGCCCCCTTCAGAACCTTAACTGTTAGTAGCAGTGGCT
GTAACAACACAAACCAGTGACCAGAGATAACAGCTTTTAGGCCAAGCTGGCCTGACGGTATGGCTGCAGG
AAGTGACTGAGCAGTAGCGGTACTCAGCCAGACCAAGACGGAGAGGGAAGAGTCCACAGCTTTCTGGAAG
CTAAGGCATTCTGGTGGTAGAAAAGTGTGCCCCAAGCCTTCATGGACGAGTTATAGGTCTTAAGATTAGT
CTCCTCTTGTTTGGATTCCATACTTGCTAAATAACCTGATAATAACCTGGTTTTCCATGTAACTGCCTCT
AGGAAGAAAATGTACTGTTCATGCTGACACAGATATTTCAGTCTGCATGGTAAAAGTTCTAAATCTTACT
ACAAAATAATAAACTGGCTGGTTTATAATGTG

>DEX0356_044.nt.2
CTCAGTGGCCACCTCCTTCTCAAGCAAAGTGGGCCCACGCAGCCAGAGAGGACAGCCTTCCTGAGTAATC
CTCAGCCCCTGATTTTGCAAACCTGAAGCACTATCAAAAACAGCAGAGTCTTCCAAGTTTATGCAGCACT
TCTGACCCAGACACACCTCTTGGGGCCCCGAGCAGCTCTGATTCAGACAGCTCCAGCAGTGAGGAGGAGG
AAGAGAAGACATCTAAGTCTGCAGTTAAGAAGAAGCCACAGAAGGTAGCAGGAGGTGCAGCCCCTTCCAA
GCCAGCCTCTGCAAAGAAAGGAAAGGCTGAGAGCAGCAACAGTTCTTCTTCTGATGACTCCAGTGAGGAA
GAGGAAGAGAAGCTCAAGGGCAAGGGCTCTCCAAGACCACAAGCCCCCAAGGCCAATGGCACCTCTGCAC
TGACTGCCCAGAATGGAAAAGCAGCTAAGAACAGTGAGGAGGAGGAAGAAGAAAAGAAAAAGGCGGCAGT
GGTAGTTTCCAAATCAGGTTCATTAAAGAAGCGGAAGCAGAATGAGGCTGCCAAGGAGGCAGAGACTCCT
CAGGCCAAGAAGATAAAGCTTCAGACCCCTAACACATTTCCAAAAAGGAAGAAAGGAGAAAAAAGGGCAT

CATCCCCATTCCGAAGGGTCAGGGAGGAGGAAATTGAGGTGGATTCACGAGTTGCGGACAACTCCTTTGA
TGCCAAGCGAGGTGCAGCCGGAGACTGGGGAGAGCGAGCCAATCAGGTTTTGAAGTTCACCAAAGGCAAG
TCCTTTCGGCATGAGAAAACCAAGAAGAAGCGGGGCAGCTACCGGGGAGGCTCAATCTCTGTCCAGGTCA
ATTCTATTAAGTTTGACAGCGAGTGACCTGAGGCCATCTTCGGTGAAGCAAGGGTGATGATCGGAGACTA
CTTACTTTCTCCAGTGGACCTGGGAACCCTCAGGTCTCTAGGTGAGGGTCTTGATGAGGACAGAAGTTTA
GAGTAGGTCCTAAGACTTTACAGTGTAACATCCTCTCTGGTCCTTTTCTGTGTTCCTAGTTTTGTACAGA
CTTGTTTTTGAGTGTTGAGTAGCAGGGACAAAATAAGGGAATGTTATTTTTTAAGAAAATTCATTTTCAT
TGTTGTCTCCTTCCTTTTCTGTGAAAGTCCTCATACTGAGAAATTTGTATATTTTATATTAAATCACTTA
CTATTGATTTTTGTTGTGATTTTCAAAGGTGGATTCCCACAGATAAAATCTTGGCTATTGCCCAAAACAT
AGTAAAGGGTCACGTGTGACTTTTTATAATAGGAAGAAAATTCTGCCTTTGTGAGTGCACATGTCCACAT
TTCATCCCTCCTTCCCTCAAAACCCTAGAGAGGGGCATTAAAGAATTGTTGATGTATATGCAATGTCTGT
TAAGCATGCACTATGTATTTCATCCTCATTTATTGGGTCTGGGACTGAAGTTTTTAGCCAGCATGGACCT
AACCTACTTTTTGGGATAAAATTCTCTGTTTTGTTACAGGCAAAATTCTGGTATGGCGTGAATGCCATGG
GTCATTCTGAATATATTTTTTTCTGTAATTTTATCATTACACGATGTTTGCAATACGTGCTTTGTTTTTT
AATTTGAAAGCAAACTTTTCTACTGTTGAAAGACATTTTTTGACAACTTGACCCTTCCTAGTATTGAGTT
CTAAGTTGAGGACTGCATCTTCTCGTTTTTTACAGTATAGAGAACAAAATGACATGAGTTTGAAAAATAC
ATATCACTTGGTATTGCTGTCTTGGTTGCAGTGGTGATACAGAATTGGTTTCATTAATTCCTACATGGTT
GAGAATCACTGATCAAGAAAGTGGGGGGAAAAAAAACAAACGTTAAAACCTCAATCCTCAGTAGGAAGGT
AGATTACATTAGGTGAAATTATAGGTAATCTATGTATGTGCTAATGGGGTTGGAAAGAACCTTACAGAGC
ATATTACCTGATAAACTGGAGTGGGTTTGGGAGAACAAACTAATAGGATTATTGTGTCTCCTAGTTGGTA
CCTGGGAGCAATTGACATGCCCCCTTCAGAACCTTAACTGTTAGTAGCAGTGGCTGTAACAACACAAACC
AGTGACCAGAGATAACAGCTTTTAGGCCAAGCTGGCCTGACGGTATGGCTGCAGGAAGTGACTGAGCAGT
AGCGGTACTCAGCCAGACCAAGACGGAGAGGGAAGAGTCCACAGCTTTCTGGAAGCTAAGGCATTCTGGT
GGTAGAAAAGTGTGCCCCAAGCCTTCATGGACGAGTTATAGGTCTTAAGATTAGTCTCCTCTTGTTTGGA
TTCCATACTTGCTAAATAACCTGATAATAACCTGGTTTTCCATGTAACTGCCTCTAGGAAGAAAATGTAC
TGTTCATGCTGACACAGATATTTCAGTCTGCATGGTAAAAGTTCTAAATCTTACTACAAAATAATAAACT
GGCTGGTTTATAATGTG

>DEX0356_045.nt.1
CGAGCGGCGCCAGTGTGATGGATAGCGTGGTCGCGGCGAGGTACAAAGATTCAGCCCAGAGGGATGACAT
GATTTTTGAAGATTGTGGCGATGTGCCCAGTGAACCGAAAGAACGAGGAGACTTTAATCATGAACGAGGA
GAGCGAAAGGTCTGCAGATTCAAGCTTGAATGGCTGGGAAATTGCTCTGGATTAAATGATGAAACTTATG
GCTACAAAGAGGGCAAACCGTGCATTATTATAAAGCTCAACCGAGTTCTAGGCTTCAAACCTAAGCCTCC
CAAGAATGAGTCCTTGGAGACTTACCCAGTGATGAAGTATAACCCAAATGTCCTTCCCGTTCAGTGCACT
GGCAAGCGAGATGAAGATAAGGATAAAGTTGGAAATGTGGAGTATTTTGGACTGGGCAACTCCCCTGGTT
TTCCTCTGCAGTATTATCCGTACTATGGCAAACTCCTGCAGCCCAAATACCTGCAGCCCCTGCTGGCCGT
ACAGTTCACCAATCTTACCATGGACACTGAAATTCGCATAGAGTGTAAGGCGTACGGTGAGAACATTGGG
TACAGTGAGAAAGACCGTTTTCAGGGACGTTTTGATGTAAAAATTGAAGTTAAGAGCTGATCACAAGCAC
AAATCTTTCCCACTAGCCATTTAATAAGTTAAAAAAAGATACAAAAACAAAAACCTACTAGTCTTGAACA
AACTGTCATACGTATGGGACCTACACTTAATCTATATGCTTTACACTAGCTTTCTGCATTTAATAGGTTA
GAATGTAAATTAAAGTGTAGCAATAGCAACAAAATATTTATTCTACTGTAAATGACAAAAGAAAAAGAAA
AATTGAGCCTTGGGACGTGCCCATTTTTACTGTAAATTATGATTCCGTAACTGACTTGTAGTAAGCAGTG
TTTCTGGCCCCTAAGTATTGCTGCCTTGTGTATTTTATTTAGTGTACAGTACTACAGGTGCATACTCTGG
TCATTTTTCAAGCCATGTTTTATTGTATCTGTTTTCTACTTTATGTGAGCAAGGTTTGCTGTCCAAGGTG
TAAATATTCAACGGGAATAAAACTGGCATGGTAATTTTTTTTTTTTGTTTGTTTTTTGTTTTTTGGCTCT
TTCAAAGGTAATGGCCCATCGATGAGCATTTTTAACATACTCCATAGTCTTTTCCTGTGGTGTTAGGTCT
TTATTTTTATTTTTTTTCCTGGGGGCTGGGGTGGGGGTTTGTCATGGGGGAACTGCCCTTTAAATTTTAAG
TGACACTACAGAAAAACACAAAAAGGTGATGGGTTGTGTTATGCTTGTATTGAATGCTGTCTTGACATCT
CTTGCCTTGTCCTCCGGTATGTTCTAAAGCTGTGTCTGAGATCTGGATCTGCCCATCACTTTGGCTAGTG
ACAGGGCTAATTAATTTGCTTTATACATTTTCTTTTACTTTCCTTTTTTCCTTTCTGGAGGCATCACATG
CTGGTGCTGTGTCTTTATGAATGTTTTAACCATTTTCATGGTGGAAGAATTTTATATTTATGCAGTTGTA
CAATTTTATTTTTTTCTGCAAGAAAAAGTGTAATGTATGAAATAAACCAAAGTCACTTGTTTGAAAATAA
ATCTTTATTTTGAACTTTATAAAAAGCAATGCAGTACCCCATAGACTGGTGTTAAATGTTGTCTACAGTG
CAAAATCCATGTTCTAACATATGTAATAATTGCCAGGAGTACAGTGCTCTTGTTGATCTTGTATTCAGTC
AGGTTAAAACAACGGACAATAAAAGAATGAACACAAAAAAAAAAAAAAAAAAAAAANNNNCCCNNNNNNN
NNNCCCCCCCCCCCCCCCCCAAAAAAAAAAAAAAAAAAAANTAANTAAAAATNTTTTTTTTTTTTTTTTT
TTTTTTTNTTTTATTNTNTAATNTNTTTTTTTTTTTTTTNTNTATTNNTTTTATAATAATAATATTTATTT
ATAATATANATTTATTTATTTTAANTATATATTAANAANAATATTTATTATATTTTTTTATAAAATTTAA
ATTTTTTAATTGTATATNNAAAAAAAATATATNNTTATTTANTAAATTTTATTTATATATTTTNTNAATN
TNAATTATTNTTTNTTTTTNTATAANATNTTTATNTTTATTATATTTTTTATTTATATAAAAATTTTATTT

TTTTTTTTTTTTTTATTTTTTTTTTAATTTTTATATTTTTTATTNTATAATTTTATATTTT

>DEX0356_046.nt.1
GCGCTCGGGCCAATGAGAGATAGGCTCATGGCGCGGCTGTAGCTGGAACTGACCAATAGGAGGGCGGGCT
GGGGGCGGAGGCCCAGGTTCCAAACGCTCCGCGGCGCCATGGGCTGAAAACTCAACCGAGATGGAATCTC
CCAGTCTGAACCGGTTTCAACCGGTTCCGAGTTTGAGGCACTAGGAGGAGGGGGAGAAGCGGCTGCAGCG
GCCGCGGCAGGAGCAGCGGGAGCTACAGCATCAGCAAGAGCAACAGTAGCTACAGCCCCGGCGGCGGTGC
CTGTTCCAGTCTTTGCTGCTGCAGTCCGTGCAACCACCCAGAGGGGGAGGGGGGAACCACCAGTCGCTGA
GGAACAAGAGAAGGGGGGAAAGTTTAGGCGAGCCTTGGGGGGGGGGGGCCAGCGCCGGAGCCGCGTGAGA
GAGGGAGCCGTGTTTTGGTAGGGGGGAGTCGGACTGCAACTGGCAGCAGAGCGTCTCCCCGGCCGTGTGG
ACTCTACACCTCCTACTCCTGCCGCTTCTGCTGCTGCCTGTGGCTGGAGGGTCCCCCTGGGGCTGAATCT
TTGGGACTTGACCCCGTTCCCTCCCCCTTCCCTCACTCCCCAGCCGGGCGGGAGCATTTATTCCCCAGAT
TAATTCCCCTTTTGGGGGGGGGCGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTTGGGGGAAGCG
TCCCTGAAATAGTAAATATTATTGAGCTCTTTTTGCCCTTTTCCTGTCCGTTTTTTTAATTTCCTTTTTT
GAGGTGGGAAAACTGAAACCCACCTTGATTCGTCCCCTCTCCCCCCTCCCCACCTTCCCTCGCCCTAATC
CCCCAACGAGGAAGGAAGGAGCAGTTGGTTCAATCTCTGGTAATCTATGCCAGCAATTATGACAATGTTA
GCAGACCATGCAGCTCGTCAGCTGCTTGATTTCAGCCAAAAACTGGATATCAACTTATTAGATAATGTGG
TGAATTGCTTATACCATGGAGAAGGAGCCCAGCAAAGAATGGCTCAAGAAGTACTGACACATTTAAAGGA
GCATCCTGATGCTTGGACAAGAGTCGACACAATTTTGGAATTTTCTCAGAATATGAATACGAAAGTAAGC
AAGTGGTGGTTTTAAAATCTGTTATTTCTTTTCTTCCAGTCATAAACTTTGAGGAAGGTATCTCATTTTT
AGTATTAGCATTTGTTATCTCAAATTATTTTATATTGTTAATTTGTTTCAAAATTTAGTCTTAAGCATAT
AATTAATTCTTGCAAATAGTCCTAAGAATTATGAATGAAGTTTTGTAATGGGACTAGTATGCAAGGTGGC
TTTTTTGGTTGTCTGATTTTGTTTTTTTCTTTTTAGTATATGTATACAGGTCTAAGTAGCATAAATTCTG
AATGTTATTCAGAGATTAAACATAACATTAAGGAATTATTTTGGGCAAGTAATCAATATAATACAGGCCA
GGCACCTCTGACACCAAGTTGTGTGGACAGCTTAAAACCCTACTCCATGGCATTGGGCTTCTAATGGGAC
AGCACTCAGTTTTTATTTACAATGGAAAATGTTATAATTCTGGTCCTTTTTTAAGTTTGAACAGAAGGGT
TGATCAAAATGTGTTTTGTCTGTTTTAGGCTTGATGATTCACGCTTCTCTTTAAACTGCCTTAAAGTAAT
AAATACTATGGCATTCTGTTTAATACACGAAAGGTTTCCACTTGATATACATTGTGTTAAAACAGAAGAT
AGCCATATTCACATCTGTTTCTGAAACCAGACTTTGTTAATTAACAGTTTTAAAATATTAATTAGTTTTG
ACATCTTGTAAATTTTAGGTCAAAAATACAGTCCAGAACTTCCTCAAAGTGCTAACCTTAAGTAACATAA
CAGTGTTTGCCTGTTTGCTTGCCTGGTAAATTTTGCTTTTTATACTGACAGACATTACTAATGTTTAAGT
TGATTTATCAACAGTATTTTTAAAAACTCAAATTTGTCTTGTGAAAAATGAGATGATTAGAGGAACAGAA
AGAAAGCAACAGGTAAAGAATTCTTCAATTTACATAAATTACAAATGATTTGGTGCTAAATTTATGCACT
ATGACATTTGGCTTTTATGCAATCATTTTTTTTTCTTTAAAGATGTTAACACTACCTGATATTAATCATA
ATGTTCCATTATGTTGATTGCTTGTAATAGGAAGAAAAGGGGTGTCTTCCTGTGCAACTGACACAGCTAG
GCTTTGACGGCAGGCCTCCACAAGTACTATGGACTACAAATTTTGGAAAATGTGATAAAAACAAGGTGGA
AGATTCTTCCAAGGAACCAGTGCGAAGGAATAAAAAAATACGTTGTTGGCCTCATTATCAAGACGTCATC
TGACCCAACTTGTGTAGAGAAAGAAAAGGTGTATATCGGAAAATTAAATATGATCCTTGTTCAGATACTG
AAACAAGAATGGCCCAAACATTGGCCAACTTTTATCAGTGATATTGTTGGAGCAAGTAGGACCAGCGAAA
GTCTCTGTCAAAATAATATGGTGATTCTTAAACTCTTGAGTGAAGAAGTATTTGATTTCTCTAGTGGACA
GATAACCCAAGTCAAATCTAAGCATTTAAAAGACAGCATGTGCAATGAATTCTCACAGATATTTCAACTG
TGTCAGTTTGTAATGGAAAATTCTCAAAATGCTCCACTTGTACATGCAACCTTGGAAACATTGCTCAGAT
TTCTGAACTGGATTCCCCTGGGATATATTTTTGAGACCAAATTAATCAGCACATTGATTTATAAGTTCCT
GAATGTTCCAATGTTTCGAAATGTCTCTCTGAAGTGCCTCACTGAGATTGCTGGTGTGAGTGTAAGCCAA
TATGAAGAACAATTTGTAACACTATTTACTCTGACAATGATGCAACTAAAGCAGATGCTTCCTTTAAATA
CCAATATTCGACTTGCGTACTCAAATGGAAAAGATGATGAACAGAACTTCATTCAAAATCTCAGTTTGTT
TCTCTGCACCTTTCTTAAGGAACATGATCAACTTATAGAAAAAAGATTAAATCTCAGGGAAACTCTTATG
GAGGCCCTTCATTATATGTTGTTGGTATCTGAAGTAGAAGAAACTGAAATCTTTAAAATTTGTCTTGAAT
ACTGGAATCATTTGGCTGCTGAACTCTATAGAGAGAGTCCATTCTCTACATCTGCCTCTCCGTTGCTTTC
TGGAAGTCAACATTTTGATGTTCCTCCCAGGAGACAGCTATATTTGCCCATGTTATTCAAGGTAACAGAG
CGGTTGGTTGAGTGTTCTTCCTGTTGCATACTGTGGTTTTGAGGTCTGAATCCAAATACTTCTAATCTGT
GTAAATAAATTAGCTATAAAAAGAGAACCCAACAACTTCTCCATGAGTGTGGAAAACTAGAACATGAAAG
GAGTTGAGTCTAGAACCTTGATTCTCAAGAGTGTGGTCCTTCTCTCAGTATCAACATTGGTTGTGATTTC
GTTAGGCAAATTCATTGGCCACCTGCCAATCTACTAAACCAGAGTCTAGGAATGAGACACAGGAAACTCC
TGTAACAGAAGTTGGTTAAAAAAATCACATTAAAACACACTTAAATAATTATAAAGCCATTTTTGTAGAA
TTACAGTGAAAAAAAATTTTTTCTTTTGGAGACAGGGTCTTGCTCTGTGGCTCAGGTTGGAGTGCAGTGG
CGTGGTCATAGCTCACTACAATCTTGAACTCCTAGACTAAAGCAATTCTCTTGTGTCAGCCTCCCAAGTA
ATTGGTTGCAGGCACACACCACCATGCCTAGCAAATTTTAAATTTTTTAGTTTTACTTCATTTATATTTA
CATAGTTTTTTCTGTTGTGCTTTAAGCCCCAGTATTTTTATTGTTTTAGAGATTGGGTGTCATTTTGTGT
TGCTCAAGCTGCTCTCAAATTCCTGGCCACAACAGTTCTTTTTGCCTCGGCCCTCCGGAATAGTTGGAAT

TATAGGCATGAATCCCTACACATGGCTGGCTTGTGGTTTTTTTGGTTTTGTTTGTTTTTAACTAAGCTTC
TACACTAATGATTTCAGTGATCTAACGATTTTTTTCTTGATTGATATATTGTTGAGTATAGGTCCGTTTA
TTAATGGTTAGTCGAATGGCTAAACCAGAGGAAGTATTGGTTGTAGAGAATGATCAAGGAGAAGTTGTGA
GAGAATTCATGAAGGATACAGATTCCATAAATTTGTATAAGAATATGAGGGAAACATTGGGTAAGTTAAT
AAATACTGTTAATCTTTACCTTTATAAATCTATTCTGGGGTGGAAATACATTTTAGGAAATGTACTTGTA
GTTTCTAAATATTTTCTAAATGTATTGTTACTCATCAATAAAATTCTGCTCATTATTTGCATTTGAAACC
ATGTTTTTCTTTGTAGTTTATCTTACTCATCTGGATTATGTAGATACAGAAAGAATAATGACAGAGAAGC
TTCACAATCAAGTGAATGGTACAGAGTGGTCATGGAAAAATTTGAATACATTGTGTTGGGCAATAGGCTC
CATTAGTGGAGCAATGCATGAAGAGGACGAAAAACGATTTCTTGTTACTGTTATAAAGGATCTATTAGGA
TTATGTGAACAGAAAAGAGGCAAAGATAATAAAGCTATTATTGCATCAAATATCATGTACATAGTAGGTC
AATACCCACGTTTTTTGAGAGCTCACTGGAAATTTCTGAAGACTGTAGTTAACAAGCTGTTCGAATTCAT
GCATGAGACCCATGATGGAGTCCAGGATATGGCTTGTGATACTTTCATTAAAATAGCCCAAAAATGCCGC
AGGCATTTCGTTCAGGTTCAGGTTGGAGAAGTGATGCCATTTATTGATGAAATTTTGAACAACATTAACA
CTATTATTTGTGATCTTCAGCCTCAACAGGTTCATACGTTTTATGAAGCTGTGGGGTACATGATTGGTGC
ACAAACAGATCAAACAGTACAAGAACACTTGATAGAAAAGTACATGTTACTCCCTAATCAAGTGTGGGAT
AGTATAATCCAGCAGGCAACCAAAAATGTGGATATACTGAAAGATCCTGAAACAGTCAAGCAGCTTGGTA
GCATTTTGAAAACAAATGTGAGAGCCTGCAAAGCTGTTGGACACCCCTTTGTAATTCAGCTTGGAAGAAT
TTATTTAGATATGCTTAATGTATACAAGTGCCTCAGTGAAAATATTTCTGCAGCTATCCAAGCTAATGGT
GAAATGGTTACAAAGCAACCATTGATTAGAAGTATGCGAACTGTAAAAAGGGAAACTTTAAAGTTAATAT
CTGGTTGGGTGAGCCGATCCAATGATCCACAGATGGTCGCTGAAAATTTTGTTCCCCCTCTGTTGGATGC
AGTTCTCATTGATTATCAGAGAAATGTCCCAGCTGCTAGAGAACCAGAAGTGCTTAGTACTATGGCCATA
ATTGTCAACAAGTTAGGGGGACATATAACAGCTGAAATACCTCAAATATTTGATGCTGTTTTTGAATGCA
CATTGAATATGATAAATAAGGACTTTGAAGAATATCCTGAACATAGAACGAACTTTTTCTTACTACTTCA
GGCTGTCAATTCTCATTGTTTCCCAGCATTCCTTGCTATTCCACCTACACAGTTTAAACTTGTTTTGGAT
TCCATCATTTGGGCTTTCAAACATACTATGAGGAATGTCGCAGATACGGGCTTACAGATACTTTTTACAC
TCTTACAAAATGTTGCACAAGAAGAAGCTGCAGCTCAGAGTTTTTATCAAACTTATTTTTGTGATATTCT
CCAGCATATCTTTTCTGTTGTGACAGACACTTCACATACTGCTGGTTTAACAATGCATGCATCAATTCTT
GCATATATGTTTAATTTGGTTGAAGAAGGAAAAATAAGTACATCATTAAATCCTGGAAATCCAGTTAACA
ACCAAATCTTTCTTCAGGAATATGTGGCTAATCTCCTTAAGTCGGCCTTCCCTCACCTACAAGATGCTCA
AGTAAAGCTCTTTGTGACAGGGCTTTTCAGCTTAAATCAAGATATTCCTGCTTTCAAGGAACATTTAAGA
GATTTCCTAGTTCAAATAAAGGAATTTGCAGGTGAAGACACTTCTGATTTGTTTTTGGAAGAGAGAGAAA
TAGCCCTACGGCAGGCTGATGAAGAGAAACATAAACGTCAAATGTCTGTCCCTGGCATCTTTAATCCACA
TGAGATTCCAGAAGAAATGTGTGATTAAAATCCAAATTCATGCTGTTTTTTTTCTCTGCAACTCGTTAGC
AGAGGAAAACAGCATGTGGGTATTTGTCGACCAAAATGATGCCAATTTGTAAATTAAAATGTCACCTAGT
GGCCCTTTTTCTTATGTGTTTTTTTGTATAAGAAATTTTCTGTGAAATATCCTTCCATTGTTTAAGCTTT
TGTTTTGGTCATCTTTATTTAGTTTGCATGAAGTTGAAAATTAAGGCATTTTTAAAAATTTTACTTCATG
CCCATTTTTGTGGCTGGGCTGGGGGGAGGAGGCAAATTCGATTTGAACATATACTTGTAATTCTAATGCA
AAATTATACAATTTTTCCTGTAAACAATACCAATTTTTAATTAGGGAGCATTTTCCTTCTAGTCTATTTC
AGCCTAGAAGAAAAGATAATGAGTAAAACAAATTGCGTTGTTTAAAGGATTATAGTGCTGCATTGTCTGA
AGTTAGCACCTCTTGGACTGAATCGTTTGTCTAGACTACATGTATTACAAAGTCTCTTTGGCAAGATTGC
AGCAAGATCATGTGCATATCATCCCATTGTAAAGCGACTTCAAAAATATGGGAACACAGTTAGTTATTTT
TACACAGTTCTTTTTGTTTTTGTGTGTGTGTGCTGTCGCTTGTCGACAACAGCTTTTTGTTTTCCTCAAT
GAGGAGTGTTGCTCATTTGTGAGCCTTCATTAACTCGAAGTGAAATGGTTAAAAATATTTATCCTGTTAG
AATAGGCTGCATCTTTTTAACAACTCATTAAAAAACAAAACAACTCTGGCTTTTGAGATGACTTATACTA
ATTTACATTGTTTACCAAGCTGTAGTGCTTTAAGAACACTACTTAAAAAGCAAAATAAACTTGGTTTACA
TTTA

>DEX0356_047.nt.1
GTGCCCGATCGCAGCAACTATTCCGGCAGCTATCATTGCTTTTGACGATAAATCGTGTCCCTCCAGTACA
ATCATAGAGATATGTAGCTGCACTACCTTACGCATTTCCAGATCCCCGATCGCATCTCTACTGTCCGTCG
CTTTACATTGCTGCTGTCCTTCCTGCTTAGACGCCTGGTGACTCTCATTTCGCAGCAGGCCACGCTGCTG
GCCTCCAATGAAGCCTTTAAAAAGCAGGCGGAGAGTGCTAGTGAGGCGGCCAAGAAGTACATGGAGGAGA
ATGACCAGCTCAAGAAGGGAGCTGCTGTTGACGGAGGCAAGTTGGATGTCGGGAATGCTGAGGTGAAGTT
GGAGGAAGAGAACAGGAGCCTGAAGGCTGACCTGCAGAAGCTAAAGGACGAGCTGGCCAGCACTAAGCAA
AAACTAGAGAAAGCTGAAAACCAGGTTCTGGCCATGCGGAAGCAGTCTGAGGGCCTCACCAAGGAGTACG
ACCGCTTGCTGGAGGAGCACGCAAAGCTGCAGGCTGCAGTAGATGGTCCCATGGACAAGAAGGAAGAGTA
AGGGCCTCCTTCCTCCCCTGCCTGCAGCTGGCTTCCACCTGGCACGTGCCTGCTGCTTCCTGAGAGCCCG
GCCTCTCCCTCCAGTACTTCTGTTTGTGCCCTTCTGCTTCCCCCATTCCCTTCCACAGCTCATAGCTCGT
CATCTCGGCCCTTGTCCACACTCTCCAAGCACATTACAGGGGACCTGATTGCTACACGTTCAGAATGCGT
TTGCTGTCATCCTGCTTGGCCTGGCCAGGCCTGGCACAGCCTTGGCTTCCACGCCTGAGCGTGGAGAGCA

CGAGTTAGTTGTAGTCCGGCTTGCGGTGGGGCTGACTTCCTGTTGGTTTGAGCCCCTTTTTGTTTTGCCC
TCTGGGTGTTTTCTTTGGTCCCGCAGGAGGGTGGGTGGAGCAGGTGGACTGGAGTTTCTCTTGAGGGCAA
TAAAAGTTGTCATGGTGTGTA

>DEX0356_047.nt.2
GCCATGGATCCCAACTGCTCCTGCGCCGCCGGTAAGAGGCTGGGGATGCCCAGTGTAGACTGTAGCGCTA
GAGAAGCAATTTCTGACCCCTCTTTCTTTCTCTGGTCACTCAATTTCAGGACAGGAGTTGCTCCTTCCCA
AAGAGTTTTGGGGTATCTTTCTCTCCATTCTAGGTTATTCGGAGCCCCCTTTTTACCGTTAAGGAGATCT
GAGTTAATGGCTTGCTCAAGTTCCCAGGAATCGGTTGTGGACTGAGGAACTCGGCCCCGGGCTCTTAGTA
CCTGCCCGGGCGGTCGGGCGGGGCCGGGGCTGCTGTGGGAGAGTTCGGTTGCTGCGGCGGGGCCTGCACG
TTGACTGTGGGAAACTCGGTGAGCGGGCTCCGCGCGCCGGGCTGGGCTCCGGGACCGCGGAGGCTCCCCG
GCCCATCGACGAGGGAGAGAGGCGAGCGGCGCGGGGAGGCCCGGGGGCCGGGGAATCTCGGGGCCCGCAG
CCTACCTGCGTGTGAGAGGCAGGGGCGAGGGGAAACCGGGCGGCGCCTCTAGAAGCTGGTCTCGGGGGCG
GGGCGGGGGGCGGTGAAGCCCCACCTGGAGGAGCGGCCGGGATGGGCCTCCGGGACGGTGTGCCAGGCCG
GGGCCAAGTCGGAGGCCCCTCGCTCTGGGTGGGCGCTGGGGCCCGCGAGGGCTACTGTAAGGACCCCTGG
CTTCTGAGGATACTGCGTCTAGAACTTTCTCCGTATGGGGCCTTGAGGTGCTTGGTCGAGACCTGCCTTT
GCGCTTGGTCCCGAATCCTGCCCTCTAGGAGTCGCTCTTGCGGGCCTCCAGCCCACCGGAGGCGAAGCGG
CCCCGGGCGGAAGGCCGCTGGATCCTCGAGGGAGGTGCCGGTTTCTCTCCGCGGGCGCCGTGGGGACGGT
GGGAGGCGGGGGCGTCGGCAGCGCTTGGACTAGGTGCGGCCTTGGGCCTGCCTGGTAGCGGGGATTTGGG
CCCGCAGAGCGCCCGCCTCTGCGGCTGAGTTCTGCCTGGCGGGGAAGGGAGCGCCCGATGGGTGCCGAGG
CGTCCTCCTCTTGGTGCCCTGGCACTGCTCTTCCCGAAGAACGCCTTTCAGTTAAACGGGCGTCGGAAAT
CTCGGGCTTCCTGGGGCAGGGATCGTCGGGAGAGGCCGCTCTGGACGTGTTGACACACGTGCTGGAGGGG
GCAGGAAACAAGCTCACATCTTCCTGTGGGAAACCTTCTAGCAACAGGATGAGTCTGCAGTGGACTGCAG
TTGCCACCTTCCTCTATGCGGAGGTCTTTGTTGTGTTGCTTCTCTGCATTCCCTTCATTTCTCCTAAAAG
ATGGCAGAAGATTTTCAAGTCCCGGCTGGTGGAGTTGTTAGTGTCCTATGGCAACACCTTCTTTGTGGTT
CTCATTGTCATCCTTGTGCTGTTGGTCATCGATGCCGTGCGCGAAATTCGGAAGTATGATGATGTGACGG
AAAAGGTGAACCTCCAGAACAATCCCGGGGCCATGGAGCACTTCCACATGAAGCTTTTCCGTGCCCAGAG
GAATCTCTACATTGCTGGCTTTTCCTTGCTGCTGTCCTTCCTGCTTAGACGCCTGGTGACTCTCATTTCG
CAGCAGGCCACGCTGCTGGCCTCCAATGAAGCCTTTAAAAAGCAGGCGGAGAGTGCTAGTGAGGCGGCCA
AGAAGTACATGGAGGAGAATGACCAGCTCAAGAAGGGAGCTGCTGTTGACGGAGGCAAGTTGGATGTCGG
GAATGCTGAGGTGAAGTTGGAGGAAGAGAACAGGAGCCTGAAGGCTGACCTGCAGAAGCTAAAGGACGAG
CTGGCCAGCACTAAGCAAAAACTAGAGAAAGCTGAAAACCAGGTTCTGGCCATGCGGAAGCAGTCTGAGG
GCCTCACCAAGGAGTACGACCGCTTGCTGGAGGAGCACGCAAAGCTGCAGGCTGCAGTAGATGGTCCCAT
GGACAAGAAGGAAGAGTAAGGGCCTCCTTCCTCCCCTGCCTGCAGCTGGCTTCCACCTGGCACGTGCCTG
CTGCTTCCTGAGAGCCCGGCCTCTCCCTCCAGTACTTCTGTTTGTGCCCTTCTGCTTCCCCCATTCCCTT
CCACAGCTCATAGCTCGTCATCTCGGCCCTTGTCCACACTCTCCAAGCACATTACAGGGGACCTGATTGC
TACACGTTCAGAATGCGTTTGCTGTCATCCTGCTTGGCCTGGCCAGGCCTGGCACAGCCTTGGCTTCCAC
GCCTGAGCGTGGAGAGCACGAGTTAGTTGTAGTCCGGCTTGCGGTGGGGCTGACTTCCTGTTGGTTTGAG
CCCCTTTTTGTTTTGCCCTCTGGGTGTTTTCTTTGGTCCCGCAGGAGGGTGGGTGGAGCAGGTGGACTGG
AGTTTCTCTTGAGGGCAATAAAAGTTGTCATGGTGTGTA

>DEX0356_047.nt.3
TCTCGCCCCTCGGCGGGGCCCCTCCCGCACCACAGAGACGTGGGACGGCCGCGCGGACTAGAGAAGCGGC
CCTCGCCGGCCCGACAAAGCCCCGCCCCCGCCCGCCGCGTGCTCGCTCCACCCCGCCCTGCTCGCCCGG
AGCCCGAGGCGCCTCTTCCTCTCCCGCTCGGAGCGCCGGCCCTCCGCTCCGGGGCGGGGTGGGAAGCGGA
GGTGGGCGGAGCCTCCGGGGCCTCGCGAGGGCTGGTGGGCGGTGCCCGGCGGGCGCTTCCGGTTTCCGGC
CGCGGTATGAGGGGCGGGGCCGGGGCTGCTGTGGGAGAGTTCGGTTGCTGCGGCGGGGCCTGCACGTTGA
CTGTGGGAAACTCGGAAACAAGCTCACATCTTCCTGTGGGAAACCTTCTAGCAACAGGATGAGTCTGCAG
TGGACTGCAGTTGCCACCTTCCTCTATGCGGAGGTCTTTGTTGTGTTGCTTCTCTGCATTCCCTTCATTT
CTCCTAAAAGATGGCAGAAGATTTTCAAGTCCCGGCTGGTGGAGTTGTTAGTGTCCTATGGCAACACCTT
CTTTGTGGTTCTCATTGTCATCCTTGTGCTGTTGGTCATCGATGCCGTGCGCGAAATTCGGAAGTATGAT
GATGTGACGGAAAAGGTGAACCTCCAGAACAATCCCGGGGCCATGGAGCACTTCCACATGAAGCTTTTCC
GTGCCCAGAGGAATCTCTACATTGCTGGCTTTTCCTTGCTGCTGTCCTTCCTGCTTAGACGCCTGGTGAC
TCTCATTTCGCAGCAGGCCACGCTGCTGGCCTCCAATGAAGCCTTTAAAAAGCAGGCGGAGAGTGCTAGT
GAGGCGGCCAAGAAGTACATGGAGGAGAATGACCAGCTCAAGAAGGGAGCTGCTGTTGACGGAGGCAAGT
TGGATGTCGGGAATGCTGAGGTGAAGTTGGAGGAAGAGAACAGGAGCCTGAAGGCTGACCTGCAGAAGCT
AAAGGACGAGCTGGCCAGCACTAAGCAAAAACTAGAGAAAGCTGAAAACCAGGTTCTGGCCATGCGGAAG
CAGTCTGAGGGCCTCACCAAGGAGTACGACCGCTTGCTGGAGGAGCACGCAAAGCTGCAGAGCCCGGCCT
CTCCCTCCAGTACTTCTGTTTGTGCCCTTCTGCTTCCCCCATTCCCTTCCACAGCTCATAGCTCGTCATC
TCGGCCCTTGTCCACACTCTCCAAGCACATTACAGGGGACCTGATTGCTACACGTTCAGAATGCGTTTGC

TGTCATCCTGCTTGGCCTGGCCAGGCCTGGCACAGCCTTGGCTTCCACGCCTGAGCGTGGAGAGCACGAG
TTAGTTGTAGTCCGGCTTGCGGTGGGGCTGACTTCCTGTTGGTTTGAGCCCCTTTTTGTTTTGCCCTCTG
GGTGTTTTCTTTGGTCCCGCAGGAGGGTGGGTGGAGCAGGTGGACTGGAGTTTCTCTTGAGGGCAATAAA
AGTTGTCATGGTGTGTA

>DEX0356_047.nt.4
TCTCGCCCCTCGGCGGGGCCCCTCCCGCACCACAGAGACGTGGGACGGCCGCGCGGACTAGAGAAGCGGC
CCTCGCCGGCCCGACAAAGCCCCGCCCCGCCCCGCCCGCGTGCTCGCTCCACCCCGCCCTGCTCGCCCGG
AGCCCGAGGCGCCTCTTCCTCTCCCGCTCGGAGCGCCGGCCCTCCGCTCCGGGGCGGGGTGGGAAGCGGA
GGTGGGCGGAGCCTCCGGGGCCTCGCGAGGGCTGGTGGGCGGTGCCCGGCGGGCGCTTCCGGTTTCCGGC
CGCGGTATGAGGGGCGGGGCCGGGGCTGCTGTGGGAGAGTTCGGTTGCTGCGGCGGGGCCTGCACGTTGA
CTGTGGGAAACTCGGAAACAAGCTCACATCTTCCTGTGGGAAACCTTCTAGCAACAGGATGAGTCTGCAG
TGGACTGCAGTTGCCACCTTCCTCTATGCGGAGGTCTTTGTTGTGTTGCTTCTCTGCATTCCCTTCATTT
CTCCTAAAAGATGGCAGAAGATTTTCAAGTCCCGGCTGGTGGAGTTGTTAGTGTCCTATGGCAACACCTT
CTTTGTGGTTCTCATTGTCATCCTTGTGCTGTTGGTCATCGATGCCGTGCGCGAAATTCGGAAGTATGAT
GATGTGACGGAAAAGGTGAACCTCCAGAACAATCCCGGGGCCATGGAGCACTTCCACATGAAGCTTTTCC
GTGCCCAGAGGAATCTCTACATTGCTGGCTTTTCCTTGCTGCTGTCCTTCCTGCTTAGACGCCTGGTGAC
TCTCATTTCGCAGCAGGCCACGCTGCTGGCCTCCAATGAAGCCTTTAAAAAGCAGGCGGAGAGTGCTAGT
GAGGCGGCCAAGAAGTACATGGAGGAGAATGACCAGCTCAAGAAGGGAGCTGCTGTTGACGGAGGCAAGT
TGGATGTCGGGAATGCTGAGGTGAAGTTGGAGGAAGAGAACAGGAGCCTGAAGGCTGACCTGCAGAAGCT
AAAGGACGAGCTGGCCAGCACTAAGCAAAAACTAGAGAAAGCTGAAAACCAGGTTCTGGCCATGCGGAAG
CAGTCTGAGGGCCTCACCAAGGAGTACGACCGCTTGCTGGAGGAGCACGCAAAGCTGCAGGGTCAGCCCT
CCCGTATTGCTGCGCCCAGGGCAGGAGAGCAGAGCAGGATGAGGCAGCCCCTGCGGGTG

>DEX0356_047.nt.5
TCTCGCCCCTCGGCGGGGCCCCTCCCGCACCACAGAGACGTGGGACGGCCGCGCGGACTAGAGAAGCGGC
CCTCGCCGGCCCGACAAAGCCCCGCCCCGCCCCGCCCGCGTGCTCGCTCCACCCCGCCCTGCTCGCCCGG
AGCCCGAGGCGCCTCTTCCTCTCCCGCTCGGAGCGCCGGCCCTCCGCTCCGGGGCGGGGTGGGAAGCGGA
GGTGGGCGGAGCCTCCGGGGCCTCGCGAGGGCTGGTGGGCGGTGCCCGGCGGGCGCTTCCGGTTTCCGGC
CGCGGTATGAGGGGCGGGGCCGGGGCTGCTGTGGGAGAGTTCGGTTGCTGCGGCGGGGCCTGCACGTTGA
CTGTGGGAAACTCGGAAACAAGCTCACATCTTCCTGTGGGAAACCTTCTAGCAACAGGATGAGTCTGCAG
TGGACTGCAGTTGCCACCTTCCTCTATGCGGAGGTCTTTGTTGTGTTGCTTCTCTGCATTCCCTTCATTT
CTCCTAAAAGATGGCAGAAGATTTTCAAGTCCCGGCTGGTGGAGTTGTTAGTGTCCTATGGCAACACCTT
CTTTGTGGTTCTCATTGTCATCCTTGTGCTGTTGGTCATCGATGCCGTGCGCGAAATTCGGAAGTATGAT
GATGTGACGGAAAAGGTGAACCTCCAGAACAATCCCGGGGCCATGGAGCACTTCCACATGAAGCTTTTCC
GTGCCCAGAGGAATCTCTACATTGCTGGCTTTTCCTTGCTGCTGTCCTTCCTGCTTAGACGCCTGGTGAC
TCTCATTTCGCAGCAGGCCACGCTGCTGGCCTCCAATGAAGCCTTTAAAAAGCAGGCGGAGAGTGCTAGT
GAGGCGGCCAAGACTCGGCCGCGACCACGCTAAGCCGAATTCCAGCAC

>DEX0356_048.nt.1
TGATTCNTTCCCCNTTTTCCTACACCCCCCCCTTCCCCCCGACCGGAGGAGCCGCTCTTTCCGCGCGGGG
CATTCNGGGGCCCGAGGNCGAGCCCGCCGCCGCCGCCGTCGCCCGAGGGAAGCGAGAAGAGGCCGCGACC
GGAGAGAAAAAGCGGAGTCGCCACCGGAGAGAAGTCGACACCCAAGCAGCAGCCGCCGCCAGANAGGCCC
GCCCACCAGNCCGCCCGTCCCCCCGCCCCGTCCACAAAGCAGCCCGCCTCCGCAAAGGGGTACGAACGAA
GGGACAAAGTCTCCAATGAAAAAAGTGTAGAAGACAGAAAGGAAGTAAAAGAAAAAAAAGAAAAAACCAA
ACAAACAATCAGTTGTCCCGGAGGAAGTGAAAAANAAAAGCATTAAAAAGAAAAAGAACATTAACACAGA
ATTGAACCAAAAGAATCCAAGCAAAAAAGAACGGACAGAACAAATTAAAGGGGGAAACGAAAAGGAGAAA
CCGGCCAGGCAAGGCCAAGGGAAACAAAAGAAAGGGCAAAGCCAAAAAGGCCAAGGGNCGACATCAAAAA
AGGGAAAGACWGGGAAGAWGAWTCAGATGAAGACATGTCTAATTTTGATCGTTTCTCTGAGATGATGAAC
AACATGGGTGGTGATGAGGATGTAGATTTACCAGAAGTAGATGGAGCAGATGATGATTCACAAGACAGTG
ATGATGAAAAAATGCCAGATCTGGAGTAAGGAATATTGTCATCACCTGGATTTTGAGAAAGAAAAATAAC
TTCTCTGCAAGATTTCATAATTGAGAGAATTCCTGAGTTGATAGCTCTAAAGGCAGATATGCTGTATTTG
CCTACTTTAACCCATTTTTCAACCTGTTTGTTTTTTAAAAGGCTTCACTAAGGGTTGATATGTACCATTG
TATGGGGCAATTTTAAGTCAGCTAAAGGCAATAACCTTTATTGCATGAAACATTTCCCCCAGACTTTTTC
ATGAAAGCTTGTTGAGGGTTCCCTAGGGGCACAATTCAATTGCAGCCAGGTTTGCGGAATTAAAATTTAA
AAAACCATTTCTTCAAACCTTAAAGGTTCCTTCCCTTTTCCTTTTCCATTTAAACCCTTTAGGAATTACT
TGGACCTTGGATTTCGGGAGAAGGCATCCTCTCCAGGCATCTTAAGGGCCGCACAGGACACGCACACTCT
CACTTTTTTAGGAAATCTTCATTCTGGAACCACCTTGGGTTTTTCCCCAACAAGGTTGGCTCCTGGGGGA
CCCCAACTGTTGGGGTGTGTTGACGACACCCCTTTTTAGCTGGGCGTGAACACCGTGGGGGAAAAGTCTC
CTCCCGCGAGGGGTGCTGATCCCCTTTCTCCGGGGTAGTGGGGGTTTCCCCCGCTTTTGAAACAGTCTTC

CTCTTGTCTTCTTCCCACAAGAGACCATTCTTTTCCCTGTCTCTGCCCACTGTGTTAGTCCTCCTCTTGT
GGGGGTCCGCGGGGGGTCGGTGAGCGTGCGTGCTCCCGGTGGGGTGTTCCCCGACGGGGGGCGTCTTCTT
TACGGGCTCGTACACCTCGCTCGAGGGAAATGTGTGCGCCCTTAAACGTTTCTGGACGGAGGGGGGCTTG
GGAGGAAAGCCTTTTTCTACACACGGGGGGGTCTCTCTGTTGGGAC

>DEX0356_049.nt.1
CAAATGGACATGTCTCTTGACTGGACATCATTAAACTGAACCGGAGCCAGCGAGGCGGCTGGGGCGGGTG
CCGGGGCCTGCGTGCCGTGTGCCGGCTCCCAGTTGCGGSCCGCGGYGGTGGGGCGCAGCCGCCGRAGTGA
ATCGAGGCGGCGGCCCATCCGGAACCGGCCGKSCATCGCCCGCGGCGCCGGGAKKGCGGCAGGAACCGAC
CGGCGCCCTACAGCAGGCCAAAACAACTTCCCGACAAGTGGCAGCACGATCTTTTCGACAGTGGCTTCGG
CGGTGGTGCCGGCGTGGAGACAGGTGGGAAACCGTGCGTGGTGTCCATCTGGATWTTGGAGTCTCAGACG
CCGATATTCAGGAACTCTTTGCTGAATTTGGAACGCTGAAGAAGGCGGCTGTGCACTATGATCGCTCTGG
TCGCAGCTTAGGAACAGCAGACGTGCACTTTGAGCGGAAGGCAGATGCCCTGAAGGCCATGAAGCAGTAC
AACGGCGTCCCTCTGGATGGCCGCCCCATGAACATTCAGCTTGTCACGTCACAGATTGACGCACAGCGGA
GGCCTGCACAGAGCGTAAACAGAGGTGGCATGACTAGAAACCGTGGCGCTGGAGGTATTGGTGGTGATGG
AGGCACCCGGAGAGGCAATGCTCGTATCTTGTGTGGAAGAGGCAGAGGTGCCGGCAGGAATTCAAAGCAG
CAGCTTTCGGCAGAGGAGCTGGATGCCCAGCTGGACGCCTATAATGCGAGAATGGACACCAGTTAAACAG
ACYRGCRWATCTYGTGYGGAACAGGACCCAGGCGTCTCCTCTTGCTCCCTGGTAAGGCGGGGCGTGGTGG
CTGTGGCTGTGCGGCCAATGATGGATTTGTTTCTTTGATGTTATTAAAATATGGATTTAAGAAACTCATG
TAACAGGTTCTTCTCTATTTATTTGCTTCTAATATATTTTTTATAGCTTCTCGAAACCCACACCTGTGTT
GGTTGACCCCGAGTGTTCCTTATTGTCGGGCGATTCAACACTTGTACACCTGGGTAGTGCGAGTGATGGT
TGCGTGAGCACAAACTGGGCGCGCGCGCGTATCGCCCGACGCATTAATCGCCCACATCAGTATTATAACT
ACTTAGGAACATACGGTCGTGTCATAGACACATTTGTGGGGAACAATTATAACCATATAGGATCTAGGAG
ACACAACCAGGGTGGGCGCGCGCGCCAGGAGCTAGATAGAAATTAGATCTCCCCGGACAGTAGCGGGGAC
GACCAGGCACCTGTGCACGACGAGAGTACGCCGAAGGTTCTTGTAATTGGTGAAAAGTACATACAAGTTA
TGGGGCGCCAATAGGTAATAGAGTGGGAGTCGCAGATAGAATATGTAGAAGGCCCAAGCGAGCGGGATGC
GAGGCGCGGGGTGTATGAATGCTTCACCCAAGGGTCGTAGGTTTAGTG

>DEX0356_050.nt.1
TGGCGTGGGTAAGACCAGCATTGCTCGCTCCATCAGCCCGCGCCCTGAACACGAGAGTACATTCCAGCTT
CAGCGTCGGGGGCATGACATGACAGTGGCTGAGATCAAGGGCACAACAGGCAGGACACTACGTGGGCAGC
CATGCCCAGGGAAGATCATCCAAGTGTTTGAAGAAGACCAAGACGGAGAACCCCACTGATCCTCATCAGA
CGAGGTGGACAAGATCAGGCCAGAGGCTACACAGGGGGACCCAGTCAGTCAGGCAACATGCATGGAGCAT
GCATGGACCCAAGAGCAAGAATGCCAAACATTCCATGGACCAACATACACATGGACAGTGCCACGTGGAC
ATTGTCACAAGGTGCATGTTCATCTGCACGGCCAACGTCACGGACACCATCCCCGAGCCGCTGCGAGACC
GTATGGAGATGATCAACGTGTCGGGCTAAGTGGSCCAGGAGAAGCTGGCCATTGCGGAGCRGCTAACTGG
TGCCCCRGGYMTCGCGCCCWTGTGTGGCWTTGGATGAGAGCAAGGCCAAGCTGTCATCGGACGTGCTGAC
GCTGCTCATCAAGCAGTACTGCCGCGAGAGCGGTGTCCGCAACCTGCAGAAGCAAGTGGAGAAGGTGTTA
CGGAAATCGGCCTACAAGATTGTCAGCGGCGAGGCCGAGTCCGTGGAGGTGACGCCCGAGAACCTGCAGG
ACTTCGTGGGGAAGCCCGTGTTCACCGTGGAGCGCATGTATGACGTGACACCGCCCGGCGTGGTCATGGG
GCTGGCCTGGACCGCAATGGGAGGCTCCACGCTGTTTGTGGAGACATCCCTGAGACGGCCACAGGACAAG
GATGCCAAGGGTGACAAGGATGGCAGCCTGGAGGTGACAGGCCAGCTGGGGGAGGTGATGAAGGAGAGCG
CCCGCATAGCCTACACCTTCGCCAGAGCCTTCCTCATGCAGCACGCCCCCGCCAATGACTACCTGGTGAC
CTCACACATCCACCTGCATGTGCCCGAGGGCGCCACCCCCAAGGACGGCCCAAGCGCAGGCTGCACCATC
GTCACGGCCCTGCTGTCCCTGGCCATGGGCAGGCCTGTCCGGCAGAATCTGGCCATGACTGGCGAAGTCT
CCCTCACGGGCAAGATCCTGCCTGTTGGTGGCATCAAGGAGAAGACCATTGCGGCCAAGCGCGCAGGGGT
GACGTGCATCGTCCTGCCAGCCGAGAACAAGAAGGACTTCTACGACCTGGCAGCCTTCATCACCGAGGGC
CTGGAGGTGCACTTCGTGGAACACTACCGGGAGATCTTCGACATCGCCTTCCCGGACGAGCAGGCAGAGG
CGCTGGCCGTGGAACGGTGACGGCCACCCCGGGACTGCAGGCGGCGGATGTCAGGCCCTGTCTGGGCCAG
AACTGGATGTCAGGCCCTGTCTGGGCCAGAACTGAGCGCTGTGGGGAGCGCGCCCGGACCTGGCAGTGGA
GCCACCGAGCGAGCAGCTCGGTCCAGTGACCCAGATCCCAGGGACCTCAGTCGGCTTAATCAGAGTGTGG
CATAGAAGCTATTTAATGATTAAAGTCATTTGCAGTAACACACACAACCACAAA

>DEX0356_051.nt.1
CGCCTTCTGTGTCCGGGTGGACATGTCTCTGGCAATTCTAGAAGAGGTAATAATGTGGGATCCTAGGGTG
TCCCGGCCGTGCCTCCCATTGGGGATGTGGCCACCCAGTACTTTGCTGACAGGGACATGTTCTGTGCTGG
CCGAGTACCTGAGGAGGATCTGAAGAGGACAATGATGGCCTGTGGAGGCTCAATCCAGACCAGTGTGAAT
GCTCTGTCAGCAGATGTGCTTGGTCGATGCCAGTTGTTTGAAGAGACCCAGATTGGAAGCGAGAGGTACA
ATCTTTTGACTGCTGCCCCARGCCAAGACATGCACCTTCATTCTCCGTGGCGGCACCGAGCAGTTTATGG
AGGAGACAGAGCGGTCCCTGCATGATGCCATCATGATCGTCAKGAGGGCCATCAAGAATGATTCAGTGGT

RGCTGGTGGCGGGGCCATTRAGATGGAACTCTCCAAGTACCTGCGGGATTACTCAAGGACTATTCCAGGA
AAACAGCAGCTGTTGATTGGGGCTTATGCCAAGGCCTTGGAGATTATCCCACGCCAGCTGTGTGACAATG
CTGGCTTTGATGCCACAAACATTCTCAACAAGCTGCGGGCTCGGCATGCCCAGGGGGGTACATGGTATGG
AGTAGACATCAACAACGAGGACATTGCTGACAACTTTGAAGCTTTCGTGTGGGAGCCAGCTATGGTGCGG
ATCAATGCGCTGACAGCAGCCTCTGAGGCTGCAGTGCCTGATCGTGTCTTGTAGATGAAACCATCAAGAA
CCTCDCCGTCTCGACTGTGGATGCTCCCACATGCATGCAGGACCGGGGTCTGTGGTCTGTGGCCGCCCCC
ATCTGAGAGGCACCCCACCCATCACATGGCTGGCTGGCTGCTGGGTGCACTTACCCTCCTTGGCTTGGTT
ACTTCATTTTACAAGGAAGGGGTAGTAATTGGCCCACTCTCTTCTTACTGGAGGCTATTTAAATAAAAGG
TAAGACTTCCA

>DEX0356_052.nt.1
GTCACCCTTTTCCTGCACAACAACGNGGTCTTCCATGATCCCCACATCGCTGACTCCTGGCCGCAGCTCC
AGGACCACTTCTCAGCTGTGRAGTCGTGGGGCCGGAGGAGGCTCTGAGCCCAGGCGAGGCCAGGGACATG
GCCATGGACCTGTGTCGGCAGGACCCCGAGTGTGAGTTCTACTTNAGCNTGGACGCNGACGCTGTNCTCA
CCAACNTGCAGACCCTGNGTNTNCTCATTGNGGAGNNNAGGGGGGTGATNGCTCCCATGNTGTNNNGGNN
GGGCAAGNTGTGGTCNAACTTNTGGGGCGNCCTGAGCNCCGATGAGTACTACGCCCGCTCNGNGGACTAN
GTGGAGCTGGTGNAGCGGAAGCGATGTNGGTGTGTGGAATGTTCCATATATCTNNNAGGCNTATGTGNTG
GGGGGTGATACCNTGNGGATGGAGCTGCCCCAGAGGGATGTGTTCTCGGGCAGTGACNCAGACNCGGACA
TGGCCTTCTGTTAGNGCTTTNGAGANAAGGGCATCTTCNTCCATCTGAGCTATCAGCATGAATTTGGCCG
GCTCCTGGCCACTTCCAGATACGACACGGAGCACCTGCACCCCGACCTCTGGCAGATCTTCGACAACCCC
GTCGACTGGAAGGAGCAGTACATCCACGAGAACTACAGCCGGGCCCTGGAAGGGAAGGAATCGTGGAGCA
GCCATGCCCGGACGTGTACTGGTTCCCACTGCTGTCAGAACAAATGTGTGATGAGCTGGTGGCAGAGATG
GAGCACTACGGCCAGTGGTCAGGCGGCCGGCATGAGGATTCAAGGCTGGCTGGAGGCTACGAGAATGTGC
CCACCGTGGACATNCCACATGAAGCAGGTGGGGTACGAGGACCAGTGGCTGCAGCTGCTGCGGACGTATG
TGGGCCCATGACCGANACCTGTTTCCGGTTTACA

>DEX0356_053.nt.1
NAAGCATAGTCTAGGTACSMSGCTKRTGMTCCCCGCCGCTGCAGCCGTCGCCTTCGGAGCGAAGGGTACC
GACCCGGCAGAAGCTCGGAGCTCTCGGGGGTATCGAGGAGGCAGGCCCGCGGGCGCACGGGCGAGCGGGCC
GGGAGCCGGAGCGGCGGAGGAGCCGGCAGCAGCGGCGCGGCGGGCTCCAGGCGAGGCGGGCTCCAGGCGA
GGCGGTCGACGCTCCTGAAAACTTGCGCGCGCGCTCGCGCCACTGCGCCCGGAGCGATGAAGATGGTCGC
GCCCTGGACTCGGTTCTACTCCAACAGCTGCTGCTTGTGCTGCCATGTCCGCACCGGCACCATCCTGCTC
GGCGTCTGGTATCTGATCATCAATGCTGTGGTACTGTTGATTTTATTGAGTGCCCTGGCTGATCCGGATC
AGTATAACTTTTCAAGTTCTGAACTGGGAGGTGACTTTGAGTTCATGGATGATGCCAACATGTGCATTGC
CATTGCGATTTCTCTTCTCATGATCCTGATATGTGCTATGGCTACTTACGGAGCGTACAAGCAACGCGCA
GCCTGGATCATCCCATTCTTCTGTTACCAGATCTTTGACTTTGCCCTGAACATGTTGGTTGCAATCACTG
TGCTTATTTATCCAAACTCCATTCAGGAATACATACGGCAACTGCCTCCTAATTTTCCCTACAGAGATGA
TGTCATGTCAGTGAATCCTACCTGTTTGGTCCTTATTATTCTTCTGTTTATTAGCATTATCTTGACTTTT
AAGGGTTACTTGATTAGCTGTGTTTGGAACTGCTACCGATACATCAATGGTAGGAACTCCTCTGATGTCC
TGGTTTATGTTACCAGCAATGACACTACGGTGCTGCTACCCCCGTATGATGATGCCACTGTGAATGGTGC
TTGCCAAGGAGCACCGGCACTTACGTGTTGCCTAAGCCTTCAAGTGGGCGGACCTGAGGGCAG

>DEX0356_054.nt.1
AAGGAGCGGATGTCTAATATATGGCCTGGCTCTAGATGCTCCGAGCGGCCGCAGTGTGATGGATTCGAGC
GGCGCCCGGGCAGGTACGGTCTGCACCCTACATGGGATTCATAGAGGCCACCAGAAACCAGGCCTTTTCT
TGTCCAGAAGTTCCAACGCTTCAGCAGCGATAGCCACCACTTACAGCTCGTCCCCACACAGTGCTGCCAC
AGCAGGCTGACGCAGGTCCTCCGCCCGAGCACCTCCAGCGCTGCACGTGCGCACTTGCGTCTGGACTGCA
CGATCCG

>DEX0356_054.nt.2
GCCAGTGTGATGGATGCGTGGTCGCGGCGAGGTACGGTCTGCACCCTACATGGGATTCATAGAGGCCACC
AGAAACCAGGCCTTTTCTTGTCCAGAAGTTCCAACGCTTCAGCAGCGATAGCCACCACTTACAGCTCGTC
CCCACACAGTGCTGCCACAGCAGGCTGACGCAGGTCCTCCGCCCGAGCACCTCCAGCGCTGCACGTGCGC
ACTTGCGTCTGGACTGCACGATCCG

>DEX0356_055.nt.1
GCCGACGAGACCTCAGAAAAAGAACAGCAAGAAGCGATTGAACACATTGATGAAGTACAAAATGAAATAG
ACAGACTTAATGAACAAGCCAGTGAGGAGATTTTGAAAGTAGAACAGAAATATAACAAACTCCGCCAACC
ATTTKTTCAGAAGAGGTCAGAATTGATCGCCAAAATCCCAACATTTTTGGGGTAACAACATTTGTCAACC
ATCCACAAGATGTCTGCACTGCTCTGCGGGAGGAAGATGAAGAGGCACTGCATTATTTGACCAGAGTTGA

AGTGACAGAATTTGAAGATATTAAATCAGGTTACAGAATAGAGTTTTDATTTTGATGGAAATCCTTACTT
TGAAAATAAAGTTCTCTCCDAAGAATTTCATCTGAATGAGAGTGGTGATCCATCTTCGAAGTCCACCGAA
ATCAAATGGAAATCTGGAAAGGATTTGACGAAACGTTCGAGTCAAACGCAGAATAAAGCCAGCAGGAAGA
GGCAGCATGAGGAACCAGAGAGCTTCTTTACCTGGTTTACTGACCATTCTGATGCAGGTGCTGATGAGTT
AGGAGAGGTCATCAAAGATGATATTTGGCCAAACCCATTACAGTACTACTTGGTTCCCGATATGGATGAT
GAAGAAGGAGAAGGAGAAGAAGATGATGATGATGATGARGAGGAGGAAGGATTAGAAGATATTGCGAAGA
AGGGGATGAGGATGAAGGTKARGAAGATKAAGATKATKATGWARTGGGAGGAAGGACAGGAGGATTAAAG
ANAAAATTTCTTTANNAACCCCCAANC

>DEX0356_001.aa.1
MASGQGPGPPRQECGEPALPSASEEQVAQDTEEVFRSYVFYRHQQEQEAEGVAAPADPEMVTLPLQPSST
MGQVGRQLAIIGDDINRRYDSEFQTMLQHLQPTAENAYEYFTKIATSLFESGINWGRVVALLGFGYRLAL
HVYQHGLTGFLGQVTRFVVDFMLHHCIARWIAQRGGWVSIQGLQCPPCCWGCPSPRPHPFLGFLYLLLAS
L

>DEX0356_002.aa.1
MAPLHSSLGNKRNSISKKKKIYSCFLNLAERIRIPEPWITPPDLQEKIHIFAQKCLFLTESLKQFTEKMQ
SDMEKIQELREAQLYSVDVTLDPDTAYPSLILSDNLRQVRYSYLQQDLPDNPERFNLFPCVLGSPCFIAG
RHYWEVEVGDKAKWTIGVCEDSVCRKGGVTSAPQNGFWAVSLWYGKEYWALTSPMTALPLRTPLQRVGIF
LDYDAGEVSFYNVTERCHTFTFSHATFCGPVRPYFSLSYSGGKSAAPLIICPMSGIDGFSGHVGNHGHSM
ETSP

>DEX0356_003.aa.1
TIYGPKVAKILRASVELLDEMKFSLEKLHQGITVSDPPFDTQPRPDDSFS

>DEX0356_003.aa.2
MFDTTPHSGRSTPSSSPSLRKRLQLLPPSRPPPEPEPGTMVEKGSDSSSEKGGVPGTPSTQSLGSRNFIR
NSKKMQSWYSMLSPTYKQRNEDFRKLFSKLPEAERLIVDYSCALQREILLQGRLYLSENWICFYSNIFRW
ETTISIQLKEVTCLKKEKTAKLIPNAIQICTESEKHFFTSFGARDRCFLLIFRLWQNALLEKTLSPRELW
HLVHQCYGSELGLTSEDEDYVSPLQLNGLGTPKEVGDVIALSDITSSGAADRSQEPSPVGSRRGHVTPNL
SRASSDADHGAEEDKEEQVDSQPDASSSQTVTPVAEPPSTEPTQPDGPTTLGPLDLLPSEELLTDTSNSS
SSTGEEGEAGGPNSFASGTCKPRSVLSKRGWIRPQH

>DEX0356_003.aa.3
MFGFNKVSLFPSLSLCPSAFPSLSSPSLPGFSLLLGLCGPSPPSLPGWGRRRDAAGRGKEGGVADAAAAS
GGSGLLCGRLGRVGAAAAEGQAGALRGRPAQPSPAQPSPALPCPALPCARGAPTAPHPCSTPHPTLAGAR
QAAPHRSGNGCSSCPQAGPHLSQNQAPWWRRDQIAPQRRVGCLGPPAPRA

>DEX0356_004.aa.1
MEDPSVRCSFQLTSGRRTSAMKVTGIFLLSALALLSLSGNTGADSLGREAKCYNELNGCTKIYDPVCGTD
GNTYPNECVLCFENRKRQTSILIQKSGPC

>DEX0356_005.aa.1
ALDVKVSHEGGRSVLPQSKYSVRHHIPYPSNGELHSQYHGYYVKLNAPQHPPVDVEDGDGSSQSSSALVH
KPSANKWSPSKSVTKPVAKESKAEPKAKKSELAIPK

>DEX0356_006.aa.1
MYMYIQTTTLTVSMSLSASVSLGMLYMPKVYIIIFHPEQNVQKRKRSFKAVVTAATMQSKLIQKGNDRPN
GEVKSELCESLETNTSSTKTTYISYSNHSI

>DEX0356_007.aa.1
MVVFLSGDNLSSPSHSGIIWQCLETFWVLTNRRRMPLASSRSAMLLSILQCPGQPHKDSSSLNVNSAEVE
EPYFTGTRQEESVCPRTLVIITCPLQYSTTQRQCFLGNGFFSLQVLHFENVKDVPFGFQTVTSDVNKLSS
FYSLKLIKRLYVDKSLNLSTEFISSTKRPYAKELETVDFKDKLEETKGQINNSIKDLTDGHFENILADNS
VNDQTKILVVNAAYFVGKWMKKFPESETKECPFRVNKTDTKPVQMMNMEATFCMGNIDSINCKIIELPFQ
NKHLSMFILLPKDVEDESTGLEKIEKQLNSESLSQWTNPSTMANAKVKLSIPKFKVEKMIDPKACLENLG
LKHIFSEDTSDFSGMSETKGVALSNVIHKVCLEITEDGGDSIEVPGARILQHKDELNADHPFIYIIRHNK
TRNIIFFGKFCSP

>DEX0356_008.aa.1

MYFSQRSRSRRHSHRRYTRSRSHSHSHRRRSRSRSYTPEYRRRRSRSHSPMSNRRRHTGSRANPDPNTCL
GVFGLSLYTTERDLREVFSRYGPLSGVNVVYDQRTGRSRGFAFVYFERIDDSKEAMERANGMELDGRRIR
VDYSITKRAHTPTPGIYMGRPTHSGGGGGGGGGGGGGGGGRRRDSYYDRGYDRGYDRYEDYDYRYRRRSP
SPYYSRYRSRSRSRSYSPRRY

>DEX0356_009.aa.1
MERMWTSEEFSFKKNIMKYILITQGLLLVEVYNLDTSTFCFCCSTVHKTFSNLQGKIVEVVEVKLKFKID
SVNTIGKEAELRIFWLFMCSVLLSSHRSCVVTQDSGSTKPELTGWIDLQPIFFHYTAAFLFLKKQKSMEQ
LLQKEAPLFNWLTESMNHCRFACPPVCFSHRFYLSEWLLHLNTVAMVSEQLKTSFLHLHSKGFIFSRNKV
VLGCFNSVDHLETREQGPWITMSTLQVFGNAGKGLHSSNLLTSFLSVRPMLACNYIFLNLKTKEISDIFS
VSFFSFHSSLKSHYCITFQIMTWIEGSLFCQGLCEVNIRISTTTKLKNEAALTEMLAKTNSLDCFCHHLL
TWNSFHIYFSYEEIQCQSDHVTSKQNFIEAYPCLNYSKIKYNKVNYFALCSSIVLSYIYILICILLSKVS
ETTCWRAVAFRQHQHHLKALLNTHYWKRGENTGIMPVIPVLWETEAGGSRGAQEFKIRLGNTARPLSLQK
QKQKTNRHGSAYLWSELLGSLRQEDHLNPGVRGCSELLLGQLQSSLGSKVRPCLLKKQNQKLDCWGPPRV
SDSVGLAVGMVGLVICLSNKRPGNAPAAGPGPHFADYYSEKSTNSDVGFLLFLISEFIWGRLFNLVDPHL
LSCEVEVNATCLETVRMRNSELKYQAQRAWMLEEMLTPMVRPSLTLLSYKVRSEIQYLLRFLSGVKCCDS
KSLIL

>DEX0356_009.aa.2
MVPPRRHRGAGRPGVLSSSPPFRLRSAKFSGIALEDLRRALKTRLQMVCVFVMNRMNSQNSGFTQRRRMA
LGIVILLLVDVIWVASSELTSYVFTQYNKPFFSTFAKTSMFVLYLLGFIIWKPWRQQCTRGLRGKHAAFF
ADAEGYFAACTTDTTMNSSLSEPLYVPVKFHDLPSEKPESTNIDTEKTPKKSRVRFSNIMEIRQLPSSHA
LEAKLSRMSYPVKEQESILKTVGKLTATQVAKISFFFCFVWFLANLSYQEALSDTQVAIVNILSSTSGLF
TLILAAVFPSNSGDRFTLSKLLAVILSIGGVVLVNLAGSEKPAGRDTVGSIWSLAGAMLYAVYIVMIKRK
VDREDKLDIPMFFGFVGLFNLLLLWPGFFLLHYTGFEDFEFPNKVVLMCIIINGLIGTVLSEFLWLWGCF
LTSSLIGTLALSLTIPLSIIADMCMQKVQFSWLFFAGAIPVFFSFFIVTLLCHYNNWDPVMVGIRRIFAF
ICRKHRIQRWNLALSPKLECSGEISGHCNLYLLGSSDSPASASRVAGTTGARHHTQLIFVFLVETGFHHV
GQDGLDLLNS

>DEX0356_010.aa.1
MDGRLDGWVDGRGWPWMRSALHTQTRWERFVEHDSLQQEYMCAYLCGQKYLHLGFGAIQEEMSQKQLNQG
LSTLWILNLKMGAGLCLKALLSHLLGPWFNKALSKLRKTTTTTQNKSSFFSSFLGKKALFFRERV

>DEX0356_010.aa.2
MLLERRSVMDVVAAEVWGLTAQNLLLSCCGWMEGWMDGWMAVDGRGCAVPCIPKPGGSVLLSMTACSRNI
CVPICVDKNIYT

>DEX0356_010.aa.3
MDSPSPMDPHMAEQTFEQAIQAASRIIRNEQFAIRRFQSMPVRLLGHSPVLRNITNSQAPDGRRKSEAGS
GAASSSGEDKENVRFWKAGVGALREEEGACWGGSLACEDPPLPSWLQDGFVFKMPWKPTHPSSTHALAEW
ASRREAFAQRPSSAPDLMCLSPDRKMEVEELSPLALGRFSLTPAEGDTEEDDGFVDILESDLKVNSLVPP
GPYLPIPGFAQKKRALSPVWQDHDVIPVSEEESEVTESQPNLWPMREEQVAGATVEGGLGGRGASHPNLE
FARETGKAVLEDSGTGGVWPMPVGCDPVLANLASDDDAVPPGMESLISAPLVKTLEKEEEKDLVMYSKCQ
RLFRSPSMPCSVIRPILKRLERPQDRDTPVQNKRRRSVTPPEEQQEAEEPKARVLRSKSLCHDEIENLLD
SDHRELIGDYSKAFLLQTVDGKHQDLKYISPETMVALLTGKFSNIVDKFVIVDCRYPYEYEGGHIKTAVN
LPLERDAESFLLKSPIAPCSLDKRVILIFHCEFSSERGPRMCRFIRERDRAVNDYPSLYYPEMYILKGGY
KEFFPQHPNFCEPQDYRPMNHEAFKDELKTFRLKTRSWAGERSRRELCSRLQDQ

>DEX0356_011.aa.1
MEDGVLKEGFLVKRGHIVHNWKARWFILRQNTLVYYKLEGGRRVTPPKGRILLDGCTITCPCLEYENRPL
LIKLKTQTSTEYFLEACSREERDAWAFEITGAIHAGQPGKVQQLHSLRNSFKLPPHISLHRIVDKMHDSN
TGIRSSPNMEQGSTYKKTFLGSSLVDWLISNSFTASRLEAVTLASMLMEENFLRPVGVRSMGAIRSGDLA
EQFLDDSTALYTFAESYKKKISPKEEISLSTVELSGTVVKQGYLAKQGHKRKNWKVRRFVLRKDPAFLHY
YDPSKEENRPVGGFSLRGSLVSALEDNGVPTGKVQLQPWPEDLRGLPPAPQ

>DEX0356_012.aa.1
MSIMSYNGGAVMAMKGKNCVAIAADRRFGIQAQMVTTDFQKIFPMGDRLYIGLAGLATDVQTVAQRLKFR
LNLYELKEGRQIKPYTLMSMVANLLYEKRFGPYYTEPVIAGLDPKTFKPFICSLDLIGCPMVTDDFVVSG
TCAEQMYGMCESLWEPNMVRWWHGEGLGSELPTLGH

>DEX0356_012.aa.2
MSIMSYNGGAVMAMKGKNCVAIAADRRFGIQAQMVTTDFQKIFPMGDRLYIGLAGLATDVQTVAQRLKFR
LNLYELKEGRQIKPYTLMSMVANLLYEKRFGPYYTEPVIAGLDPKTFKPFICSLDLIGCPMVTDDFVVSG
TCAEQMYLGRDHAKPNSS

>DEX0356_013.aa.1
MLFPQVPPPRPGLADPPLPFPSSNLNKPYLFSVLLGAWQLGNPGSRSQKVGVAWVEPHPVYSWKEGACAD
IALVRLERSIQFSERVLPICLPDASIHLPPNTHCWISGWGSIQDGVPLPHPQTLQKLKVPIIDSEVCSHL
YWRGAGQGPITEDMLCAGYLEGERDACLGDSGGPLMCQVDGAWLLAGIISWGEGCAERNRPGVYISLSAH
RSWVEKIVQGVQLRGRAQGGGALRAPSQGSGAAARS

>DEX0356_014.aa.1
MRLWLGRDAEGTELSRRHNWTKPEPQAPVAWERVAPSNLPQGHPLPKSFSSPPSPSNKREEEEEEFNFEV
IPPPPEFSNDPEPPAPALQYLGRQSSPPRNNYSDLRQLPNAGPGAPPALGFSRFPAGARYAGAGGLERFS
GGGRSLIKKRLYVGEPHRGPGLPHGGTGRSLSSPNCFGPQPGGPEMRRVNSAGRAPPGGLHAPRLSLEGA
ARGAAEAKHKAPGSADYGFAPAAGRSPYTTTRYGSPINTFTVRPGTRHPISYVCSGAHRKATS

>DEX0356_015.aa.1
MAALVLEDGSVLRGQPFGAAVSTAGEVVFQTGMVGYPEALTDPSYKAQILVLTYPLIGNYGIPPDEMDEF
GLCKWFESSGIHVAALVVGECCPTPSHWSATRTLHEWLQQHGIPGLQGVDTRELTKKLREQGSLLGKLVQ
NGTEPSSLPFLDPNARPLVPEVSIKTPRVFNTGGAPRILALDCGLKYNQIRCLCQRGAEVTVVPWDHALD
SQEYEGLFLSNGPGDPASYPSVVSTLSRVLSEPNPRPVFGICLGHQLLALAIGAKTYKMRYGNRGHNQPC
LLVGSGRCFLTSQNHGFAVETDSLPADWAPLFTNANDGSNEGIVHNSLPFFSVQFHPEHQAGPSDMELLF
DIFLETVKEATAGNPGGQTVRERLTERLCPPGIPTPGSGLPPPRKVLILGSGGLSIGQAGEFDYSGSQAI
KALKEENIQTLLINPNIATVQTSQGLADKVYFLPITPHYVTQVIRNERPDGVLLTFGGQTALNCGVELTK
AGVLARYGVRVLGTTVETIELTEDRRAFAARMAEIGEHVAPSEAGNSLEQAQAAAERLGYPVLVRAAFAV
GGLGSGFASNREELSALVAPAFAHTSQVLVDKSLKGWKEIEYEVVRDAYGNCVTVCNMENLDPLGIHTGE
SIVVAPSQTLNDREYQLLRQTAIKVTQHLGIVGECNVQYALNPESEQYYIIEVNARLSRSSALASKATGY
PLAYVAAKLALGIPLPELRNSVTGGTAAFEPSVDYCVVKIPRWDLSKFLRVSTKIGSCMKSVGEVMGIGR
SFEEAFQKALRMVDENCVGFDHTVKPVSDMELETPTDKRIFVVAAALWAGYSVDRLYELTRIDRWFLHRM
KRIIAHAQLLEQHRGQPLPPDLLQQAKCLGFSDKQIALAVLSTELAVRKLRQELGICPAVKQIDTVAAEW
PAQTNYLYLTYWGTTHDLTFRTPHVLVLGSGVYRIGSSVEFDWCAVGCIQQLRKMGYKTIMVNYNPETVS
TDYDMCDRLYFDEISFEVVMDIYELENPEGVILSMGGQLPNNMAMALHRQQCRVLGTSPEAIDSAENRFK
FSRLLDTIGISQPQWRELSDLESARQFCQTVGYPCVVRPSYVLSGAAMNVAYTDGDLERFLSSAAAVSKE
HPVVISKFIQEAKEIDVDAVASDGVVAAIAISEHVENAGVHSGDATLVTPPQDITAKTLERIKAIVHAVG
QELQVTGPFNLQLIAKDDQLKVIECNVRVSRSFPFVSKTLGVDLVALATRVIMGEEVEPVGLMTGSGVVG
VKVPQFSFSRLAGADVVLGVEMTSTGEVAGFGESRCEAYLKAMLSTGFKIPKKNILLTIGSYKNKSELLP
TVRLLESLGYSLYASLGTADFYTEHGVKVTAVDWHFEEAVDGECPPQRSILEQLAEKNFELVINLSMRGA
GGRRLSSFVTKGYRTRRLAADFSVPLIIDIKCTKLFVEALGQIGPAPPLKVHVDCMTSQKLVRLPGLIDV
HVHLREPGGTHKEDFASGTAAALAGGITMVCAMPNTRPPIIDAPALALAQKLAEAGARCDFALFLGASSE
NAGTLGTVAGSAAGLKLYLNETFSELRLDSVVQWMEHFETWPSHLPIVAHAEQQTVAAVLMVAQLTQRSV
HICHVARKEEVRVHLRSCCPCCFPSNTKGQVVLRGRRQQEESLETAGGGSRASTLAGLHHHFFEQGCWPL
GLCVGQGIWVVPLLDLPIVPQILLIKAAKARGLPVTCEVAPHHLFLSHDDLERLGPGKGEVRPELGSRQD
VEALWENMAVIDCFASDHAPHTLEEKCGSRPPPGFPGLETMLPLLLTAVSEGRLSLDDLLQRLHHNPRRI
FHLPPQEDTYVEVDLEHEWTIPSHMPFSKAHWTPFEGQKVKGTVRRVVLRGEVAYIDGQVLVPPGYGQDV
RKWPQGAVPQLPPSAPATSEMTTTPERPRRGIPGLPDGRFHLPPRIHRASDPGLPAEEPKEKSSRKVAEP
ELMGTPDGTCYPPPPVPRQASPQNLGTPGLLHPQTSPLLHSLVGQHILSVQQFTKDQMSHLFNVAHTLRM
MVQKERSLDILKGKVMASMFYEVSTRTSSSFAAAMARLGGAVLSFSEATSSVQKGESLADSVQTMSCYAD
VVVLRHPQPGAVELAAKHCRRPVINAGDGVGEHPTQALLDIFTIREELGTVNGMTITMVGDLKHGRTVHS
LACLLTQYRVSLRYVAPPSLRMPPTVRAFVASRGTKQEEFESIEEALPDTDVLYMTRIQKERFGSTQEYE
ACFGQFILTPHIMTRAKKKMVVMHPMPRVNEISVEVDSDPRAAYFRQAENGMYIRMALLATVLGRF

>DEX0356_015.aa.2
MAALVLEDGSVLRGQPFGAAVSTAGEVVFQTGMVGYPEALTDPSYKAQILVLTYPLIGNYGIPPDEMDEF
GLCKWFESSGIHVAALVVGECCPTPSHWSATRTLHEWLQQHGIPGLQGVDTRELTKKLREQGSLLGKLVQ
NGTEPSSLPFLDPNARPLVPEVSIKTPRVFNTGGAPRILALDCGLKYNQIRCLCQRGAEVTVVPWDHALD
SQEYEGLFLSNGPGDPASYPSVVSTLSRVLSEPNPRPVFGICLGHQLLALAIGAKTYKMRYGNRGHNQPC
LLVGSGRCFLTSQNHGFAVETDSLPADWAPLFTNANDGSNEGIVHNSLPFFSVQFHPEHQAGPSDMELLF
DIFLETVKEATAGNPGGQTVRERLTERLCPPGIPTPGSGLPPPRKVLILGSGGLSIGQAGEFDYSGSQAI
KALKEENIQTLLINPNIATVQTSQGLADKVYFLPITPHYVTQVIRNERPDGVLLTFGGQTALNCGVELTK

AGVLARYGVRVLGTTVETIELTEDRRAFAARMAEIGEHVAPSEAGNSLEQAQAAAERLGYPVLVRAAFAV
GGLGSGFASNREELSALVAPAFAHTSQVLVDKSLKGWKEIEYEVVRDAYGNCVTVCNMENLDPLGIHTGE
SIVVAPSQTLNDREYQLLRQTAIKVTQHLGIVGECNVQYALNPESEQYYIIEVNARLSRSSALASKATGY
PLAYVAAKLALGIPLPELRNSVTGGTAAFEPSVDYCVVKIPRWDLSKFLRVSTKIGSCMKSVGEVMGIGR
SFEEAFQKALRMVDENCVGFDHTVKPVSDMELETPTDKRIFVVAAALWAGYSVDRLYELTRIDRWFLHRM
KRIIAHAQLLEQHRGQPLPPDLLQQAKCLGFSDKQIALAVLSTELAVRKLRQELGICPAVKQIDTVAAEW
PAQTNYLYLTYWGTTHDLTFRTPHVLVLGSGVYRIGSSVEFDWCAVGCIQQLRKMGYKTIMVNYNPETVS
TDYDMCDRLYFDEISFEVVMDIYELENPEGVILSMGGQLPNNMAMALHRQQCRVLGTSPEAIDSAENRFK
FSRLLDTIGISQPQWRELSDLESARQFCQTVGYPCVVRPSYVLSGAAMNVAYTDGDLERFLSSAAAVSKE
HPVVISKFIQEAKEIDVDAVASDGVVAAIAISEHVENAGVHSGDATLVTPPQDITAKTLERIKAIVHAVG
QELQVTGPFNLQLIAKDDQLKVIECNVRVSRSFPFVSKTLGVDLVALATRVIMGEEVEPVGLMTGSGVVG
VKVPQFSFSRLAGADVVLGVEMTSTGEVAGFGESRCEAYLKAMLSTGFKIPKKNILLLTIGSYKNKSELLP
TVRLLESLGYSLYASLGTADFYTEHGVKVTAVDWHFEEAVDGECPPQRSILEQLAEKNFELVINLSMRGA
GGRRLSSFVTKGYRTRRLAADFSVPLIIDIKCTKLFVEALGQIGPAPPLKVHVDCMTSQKLVRLPGLIDV
HVHLREPGGTHKEDFASGTAAALAGGITMVCAMPNTRPPIIDAPALALAQKLAEAGARCDFALFLGASSE
NAGTLGTVAGSAAGLKLYLNETFSELRLDSVVQWMEHFETWPSHLPIVAHAEQQTVAAVLMVAQLTQRSV
HICHVARKEEILLIKAAKARGLPVTCEVAPHHLFLSHDDLERLGPGKGEVRPELGSRQDVEALWENMAVI
DCFASDHAPHTLEEKCGSRPPPGFPGLETMLPLLLTAVSEGRLSLDDLLQRLHHNPRRIFHLPPQEDTYV
EVDLEHEWTIPSHMPFSKAHWTPFEGQKVKGTVRRVVLRGEVAYIDGQVLVPPGYGQDVRKWPQGAVPQL
PPSAPATSEMTTTPERPRRGIPGLPDGRFHLPPRIHRASDPGLPAVFLRPGAGIPRGSRAWAEEPKEKSS
RKVAEPELMGTPDGTCYPPPPVPRQASPQNLGTPGLLHPQTSPLLHSLVGQHILSVQQFTKDQMSHLFNV
AHTLRMMVQKERSLDILKGKVMASMFYEVSTRTSSSFAAAMARLGGAVLSFSEATSSVQKGESLADSVQT
MSCYADVVVLRHPQPGAVELAAKHCRRPVINAGDGVGEHPTQALLDIFTIREELGTVNGMTITMVGDLKH
GRTVHSLACLLTQYRVSLRYVAPPSLRMPPTVRAFVASRGTKQEEFESIEEALPDTDVLYMTRIQKERFG
STQEYEACFGQFILTPHIMTRAKKKMVVMHPMPRVNEISVEVDSDPRAAYFRQAENGMYIRMALLATVLG
RF

>DEX0356_015.aa.3
MYICPLVLFAAEEPKEKSSRKVAEPELMGTPDGTCYPPPPVPRQASPQNLGTPGLLHPQTSPLLHSLVGQ
HILSVQQFTKDQMSHLFNVAHTLRMMVQKERSLDILKGKVMASMFYEVSTRTSSSFAAAMARLGGAVLSF
SEATSSVQKGESLADSVQTMSCYADVVVLRHPQPGAVELAAKHCRRPVINAGDGVGEHPTQALLDIFTIR
EELGTVNGMTITMVGDLKHGRTVHSLACLLTQYRVSLRYVAPPSLRMPPTVRAFVASRGTKQEEFESIEE
ALPDTDVLYMTRIQKERFGSTQEYEACFGQFILTPHIMTRAKKKMVVMHPMPRVNEISVEVDSDPRAAYF
RQAENGMYIRMALLATVLGRF

>DEX0356_016.aa.1
MAGALAGLAAGLQVPRVAPSPDSDSDTDSEDPSLRRSAGGLLRSQVIHSGHFMVSSPHSDSLPRRRDQEG
SVGPSDFGPRSIDPTLTRLFECLSLAYSGKLVSPKWKNFKGLKLLCRDKIRLNNAIWRAWYIQYVKRRKS
PVCGFVTPLQGPEADAHRKPEAVVLEGNYWKRRIEVVMREYHKWRIYYKKRVSGGGPGRPQSFPPAAAGY
RPPRKIPGKGILTPELAPLGPSIQSRADSATVWPQRLLAASLPRGRLRKPSREDDLLAPKQAEGRWPPPE
QWCKQLFSSVVPVLLGDPEEEPGGRQLLDLNCFLSDISDTLFTMTQSGPSPLQLPPEDAYVGNADMIQPD
LTPLQPSLDDFMDISDFFTNSRLPQPPMPSNFPEPPSFSPVVDSLFSSGTLGPEVPPASSAMTHLSGHSR
LQARNSCPGPLDSSAFLSSDFLLPEDPKPRLPPPPVPPPLLHYPPPAKVPGLEPCPPPPFPPMAPPTALL
QEEPLFSPRFPFPTVPPAPGVSPLPAPAAFPPTPQSVPSPAPTPFPIELLPLGYSEPAFGPCFSMPRGKP
PAPSPRGQKASPPTLAPATASPPTTAGSNNPCLTQLLTAAKPEQALEPPLVSSTLLRSPGSPQETVPEFP
CTFLPPTPAPTPPRPPPGPATLAPSRPLLVPKAERLSPPAPSGKERGCRDWGNLGDEGKRRKKKGTGSGA
RWEEGGNGDPFPWVGLSVGLSVPTTGSERRLSGDLSSMPGPGTLSVRVSPPQPILSRGRPDSNKTENRRI
THISAEQKRRFNIKLGFDTLHGLVSTLSAQPSLKVSKATTLQKTAEYILMLQQERAGLQEEAQQLRDEIE
ELNAAINLCQQQLPATGVPITHQRFDQMRDMFDDYVRTRTLHNWKFWVFSILIRPLFESFNGMVSTASVH
TLRQTSLAWLDQYCSLPALRPTVLNSLRQLGTSTSILTDPGRIPEQATRAVTEGTLGKPL

>DEX0356_016.aa.2
MAVPWEQASLGGQRPCLGTQRPATDFFTNSRLPQPPMPSNFPEPPSFSPVVDSLFSSGTLGPEVPPASSA
MTHLSGHSRLQARNSCPGPLDSSAFLSSDFLLPEDPKPRLPPPPVPPPLLHYPPPAKVPGLEPCPPPPFP
PMAPPTALLQEEPLFSPRFPFPTVPPAPGVSPLPAPAAFPPTPQSVPSPAPTPFPIELLPLGYSEPAFGP
CFSMPRGKPPAPSPRGQKASPPTLAPATASPPTTAGSNNPCLTQLLTAAKPEQALEPPLVSSTLLRSPGS
PQETVPEFPCTFLPPTPAPTPPRPPPGPATLAPSRPLLVPKAERLSPPAPSGKERGCRDWGNLGDEGKRR
KKKGTGSGARGEEGGNGDPFPWVGLSVGLSVPTTGSERRLSGDLSSMPGPGTLSVRVSPPQPILSRGRPD
SNKTENRRITHISAEQKRRFNIKLGFDTLHGLVSTLSAQPSLKVSKATTLQKTAEYILMLQQERAGLQEE
AQQLRDEIEELNAAINLCQQQLPATGVPITHQRFDQMRDMFDDYVRTRTLHNWKFWVFSILIRPLFESFN

GMVSTASVHTLRQTSLAWLDQYCSLPALRPTVLNSLRQLGTSTSILTDPGRIPEQATRAVTEGTLGKPL

>DEX0356_016.aa.3
MAGALAGLAAGLQVPRVAPSPDSDSDTDSEDPSLRRSAGGLLRSQVIHSGHFMVSSPHSDSLPRRRDQEG
SVGPSDFGPRSIDPTLTRLFECLSLAYSGKLVSPKWKNFKGLKLLCRDKIRLNNAIWRAWYIQYVKRRKS
PVCGFVTPLQGPEADAHRKPEAVVLEGNYWKRRIEVVMREYHKWRIYYKKRVSGGGPGRPQSFPPAAAGY
RPPRKIPGKGILTPELAPLGPSIQSRADSATVWPQRLLAASLPRGRVSSCGKWGISSLVSRGDFGGKVME
DGAKGSGVQLCVSLQLRKPSREDDLLAPKQAEGRWPPPEQWCKQLFSSVVPVLLGDPEEEPGGRQLLDLN
CFLSDISDTLFTMTQSGPSPLQLPPEDAYVGNADMIQPDLTPLQPSLDDFMDISDFFTNSRLPQPPMPSN
FPEPPSFSPVVDSLFSSGTLGPEVPPASSAMTHLSGHSRLQARNSCPGPLDSSAFLSSDFLLPEDPKPRL
PPPPVPPPLLHYPPPAKVPGLEPCPPPPFPPMAPPTALLQEEPLFSPRFPFPTVPPAPGVSPLPAPAAFP
PTPQSVPSPAPTPFPIELLPLGYSEPAFGPCFSMPRGKPPAPSPRGQKASPPTLAPATASPPTTAGSNNP
CLTQLLTAAKPEQALEPPLVSSTLLRSPGSPQETVPEFPCTFLPPTPAPTPPRPPPGPATLAPSRPLLVP
KAERLSPPAPSGKERGCRDWGNLGDEGKRRKKKGTGSGARWEEGGNGDPFPWVGLSVGLSVPTTGSERRL
SGDLSSMPGPGTLSVRVSPPQPILSRGRPDSNKTENRRITHISAEQKRRFNIKLGFDTLHGLVSTLSAQP
SLKVSKATTLQKTAEYILMLQQERAGLQEEAQQLRDEIEELNAAINLCQQQLPATGVPITHQRFDQMRDM
FDDYVRTRTLHNWKFWVFSILIRPLFESFNGMVSTASVHTLRQTSLAWLDQYCSLPALRPTVLNSLRQLG
TSTSILTDPGRIPEQATRAVTEGTLGKPL

>DEX0356_017.aa.1
VLPCPGPGPLVTSCFPPTVSPYPHPGPGWSPAHRGATGGNGPDSGATDSPAVGDDDREQVAGPKGKAPPV
PAPARESGNRSARPLHSLSVLAFDQERLERKILALRQARRPVPPEVAQQYQDIMQRSQWQRAQLEQGGVG
IRREYAAQLERQLQFYTEAARRLGNDGSRDAAKEALYRRNLVESELQRLRR

>DEX0356_017.aa.2
EGLGDGHLEGEGLGDGHLEGEGLGDGHLEGEGLGDGHLAGDGLGDGHSPGGLWAVRPHPKPPFPRHPFSE
YAAQLERQLQFYTEAARRLGNDGSRDAAKEALYRRNLVESELQRLRR

>DEX0356_017.aa.3
EGLGDGHLEGEGLGDGHLEGEGLGDGHLEGEGLGDGHLAGDGLGDGHSPGGLWAVRPHPKPPFPRHPFSE
YAAQLERQLQFYTEAARRLGNDGSRVSWSRAGWALGSGQGGACRDYLLNAHPPQDAAKEALYRRNLVESE
LQRLRR

>DEX0356_018.aa.1
MTSSGPGPRFLLLLPLLLPPAASASDRPRGRDPVNPEKLLVITVATAETEGYLRFLRSAEFFNYTVRTLG
LGEEWRGGDVARTVGGGQKVRWLKKEMEKYADREDMIIMFVDSYDVILAGSPTELLKKFVQSGSRLLFSA
ESFCWPEWGLAEQYPEVGTGKRFLNSGGFIGFATTIHQIVRQWKYKDDDDDQLFYTRLYLDPGLREKLSL
NLDHKSRIFQNLNGALDEVVLKFDRNRVRIRNVAYDTLPIVVHGNGPTKLQLNYLGNYVPNGWTPEGGCG
FCNQDRRTLPGGQPPPRVFLAVFVEQPTPFLPRFLQRLLLLDYPPDRVTLFLHNNEVFHEPHIADSWPQL
QDHFSAVKLVGPEEALSPGEARDMAMDLCRQDPECEFYFSLDADAVLTNLQTLRILIEENRKVIAPMLSR
HGKLWSNFWGALSPDEYYARSEDYVELVQRKRVGVWNVPYISQAYVIRGDTLRMELPQRDVFSGSDTDPD
MAFCKSFRDKGIFLHLSNQHEFGRLLATSRYDTEHLHPDLWQIFDNPVDWKEQYIHENYSRALEGEGIVE
QPCPDVYWFPLLSEQMCDELVAEMEHYGQWSGGRHETALQECLMPVELCPAVSSSHHTASKAAMSSLGAI
EATGKP

>DEX0356_018.aa.2
MTSSGPGPRFLLLLPLLLPPAASASDRPRGRDPVNPEKLLVITVATAETEGYLRFLRSAEFFNYTVRTLG
LGEEWRGGDVARTVGGGQKVRWLKKEMEKYADREDMIIMFVDSYDVILAGSPTELLKKFVQSGSRLLFSA
ESFCWPEWGLAEQYPEVGTGKRFLNSGGFIGFATTIHQIVRQWKYKDDDDDQLFYTRLYLDPGLREKLSL
NLDHKSRIFQNLNGALDEVVLKFDRNRVRIRNVAYDTLPIVVHGNGPTKLQLNYLGNYVPNGWTPEGGCG
FCNQDRRTLPGGQPPPRVFLAVFVEQPTPFLPRFLQRLLLLDYPPDRVTLFLHNNEVFHEPHIADSWPQL
QDHFSAVKLVGPEEALSPGEARDMAMDLCRQDPECEFYFSLDADAVLTNLQTLRILIEENRKVIAPMLSR
HGKLWSNFWGALSPDEYYARSEDYVELVQRKRVGVWNVPYISQAYVIRGDTLRMELPQRDVFSGSDTDPD
MAFCKSFRDKGIFLHLSNQHEFGRLLATSRYDTEHLHPDLWQIFDNPVDWKEQYIHENYSRALEGEGIVE
QPCPDVYWFPLLSEQMCDELVAEMEHYGQWSGGRHEVRAWARRAEAFPEEEVQDLVNKQWNQENGSKEGF
PGGRQTLGSQLHSFMYSRNIYCWAQWLMPVILALWEAEAGGLLEQGEFEASL

>DEX0356_018.aa.3
MTSSGPGPRFLLLLPLLLPPAASASDRPRGRDPVNPEKLLVITVATAETEGYLRFLRSAEFFNYTVRTLG
LGEEWRGGDVARTVGGGQKVRWLKKEMEKYADREDMIIMFVDSYDVILAGSPTELLKKFVQSGSRLLFSA

ESFCWPEWGLAEQYPEVGTGKRFLNSGGFIGFATTIHQIVRQWKYKDDDDDQLFYTRLYLDPGLREKLSL
NLDHKSRIFQNLNGALDEVVLKFDRNRVRIRNVAYDTLPIVVHGNGPTKLQLNYLGNYVPNGWTPEGGCG
FCNQDRRTLPGGQPPPRVFLAVFVEQPTPFLPRFLQRLLLLDYPPDRVTLFLHNNEVFHEPHIADSWPQL
QDHFSAVKLVGPEEALSPGEARDMAMDLCRQDPECEFYFSLDADAVLTNLQTLRILIEENRKVIAPMLSR
HGKLWSNFWGALSPDEYYARSEDYVELVQRKRVGVWNVPYISQAYVIRGDTLRMELPQRDVFSGSDTDPD
MAFCKSFRDKGIFLHLSNQHEFGRLLATSRYDTEHLHPDLWQIFDNPVDWKEQYIHENYSRALEGEGIVE
QPCPDVYWFPLLSEQMCDELVAEMEHYGQWSGGRHEWEGTNGAGPHMKGGPGQVIGSLADNLAATGGGKG
PAKSGPNHQHRARGQTGTPKGDTG

>DEX0356_018.aa.4
MQGPYVQGLGYPEGRRTSRDELGRIWGAVLPADRGVGASRGGVWNVPYISQAYVIRGDTLRMELPQRDVF
SGSDTDPDMAFCKSFRDKVSAGARSGLGAVPQTPGIACITDTPTPLTGIFLHLSNQHEFGRLLATSRYDT
EHLHPDLWQIFDNPVVSGDPSPEHTGSPPCAKEDTKRDGCCQRRTLSGRGRSPGGWERGGHRSDVPPTLS
QDWKEQYIHENYSRALEGEGIVEQPCPDVYWFPLLSEQMCDELVAEMEHYGQWSGGRHEDSRLAGGYENV
PTVDIHMKQVGYEDQWLQLLRTYVGPMTESLFPGYHTKARAVMNFVVRYRPDEQPSLRPHHDSSTFTLNV
ALNHKGLDYEGGGCRFLRYDCVISSPRKGWALLHPGRLTHYHEGLPTTWGTRYIMVSFVDP

>DEX0356_019.aa.1
MRSPELALPRGMQPTEFFQSLGGDGERNVQIEMAHGTTTLAFKFQHGVIAAVDSRASAGSYISALRVNKV
IEINPYLLGTMSGCAADCQYWERLLAKECRLYYLRNGERISVSAASKLLSNMMCQYRGMGLSMGSMICGW
DKKGPGLYYVDEHGTRLSGNMFSTGSGNTYAYGVMDSGYRPNLSPEEAYDLGRRAIAYATHRDSYSGGVV
NMSTVSSWSGPLERPMSCPVCPSGSASIAAGAGGRPGRTCGPPGHTLSPGKKGVVCQQKVVSTSGKHMGR
YKSRVCAAKMIIIGGGSTPYPYNGSRVHTSYY

>DEX0356_019.aa.2
MRSPELALPRGMQPTEFFQSLGGDGERNVQIEMAHGTTTLAFKFQHGVIAAVDSRASAGSYISALRVNKV
IEINPYLLGTMSGCAADCQYWERLLAKECRLYYLRNGERISVSAASKLLSNMMCQYRGMGLSMGSMICGW
DKKGPGLYYVDEHGTRLSGNMFSTGSGNTYAYGVMDSGYRPNLSPEEAYDLGRRAIAYATHRDSYSGGVV
NIPEDGGDSERPFGRASRDHQGGGGRPRGGSGTPFRKSRKVSRGGCLNQWPKKGPAAYWRSSTY

>DEX0356_019.aa.3
MRSPELALPRGMQPTEFFQSLGGDGERNVQIEMAHGTTTLAFKFQHGVIAAVDSRASAGSYISALRVNKV
IEINPYLLGTMSGCAADCQYWERLLAKECRLYYLRNGERISVSAASKLLSNMMCQYRGMGLSMGSMICGW
DKKGPGLYYVDEHGTRLSGNMFSTGSGNTYAYGVMDSGYRPNLSPEEAYDLGRRAIAYATHRDSYSGGVV
NMYHIDTTMVGESKVDVDLCPPEAQSG

>DEX0356_020.aa.1
MDSSLPNAQHDVQGLSPQENLSPPSFPTSSHPQVCQKSGEISLLKQQLKESQAELVQKGSELVALRVALR
EARATLRVSEGRARGLQEAARARELELEACSQELQRHRQEAEQLREKAGQLDAEAAGLREPPVPPATADP
FLLAESDEAKVQRAAAGVGGSLRAQVERLRVELQRERRRGEEQRDSFEGERLAWQAEKEQVIRYQKQLQH
NYIQMYRRNRQLEQELQQLSLELEARELADLGLAEQAPCICLEEITATEI

>DEX0356_020.aa.2
MRGSHLPQPQTHFSKWSLHLPLLLFRAPRAGAGLAGPGFPDSWLWLPGCVALRARLCGTLDLWALGSVGL
GGPGSLGPAGCPLVGLTLGPLRVFVSFGSLSFSEPVSFLQFSLSPCSLKWGNATEVPAPLAP

>DEX0356_021.aa.1
MERHQPRLHHPAQGSAAGTPYPSSASLRGCRESKMPRRKGPQHPPPPSGPEEPGEKRPKFHLNIRTLTDD
MLDKFASIRIPGSKKERPPLPNLKTAFASSDCSAAPLEMMENFPKPLSENELLELFEKMMEDMNLNEDKK
APLREKDFSIKKEMVMQYINTASKTGSLKRSRQISPQEFIHELKMGSADERLVTCLESLRVSLTSNPVSW
VESFGHEGLGLLLDILEKLISGKIQEKVVKKNQHKVIQCLKALMNTQYGLERIMSEERSLSLLAKAVDPR
HPNMMTDVVKLLSAVCIVGEESILEEVLEALTSAGEEKKIDRFFCIVEGLRHNSVQLQVACMQLINALVT
SPDDLDFRLHIRNEFMRCGLKEILPNLKCIKNDGLDIQLKVFDEHKEEDLFELSHRLEDIRAELDEAYDV
YNMVWSTVKETRAEGYFISILQHLLLIRNDYFIRQQYFKLIDECVSQIVLHRDGMDPDFTYRKRLDLDLT
QFVDICIDQAKLEEFEEKASELYKKFEKEFTDHQETQAELQKKEAKINELQAELQAFKSQFGALPADCNI
PLPPSKEGGTGHSALPPPPPLPSGGGVPPPPPPPPPPPLPGMRMPFSGPVPPPPPLGFLGGQNSPPLPIL
PFGLKPKKEFKPEISMRRLNWLKIRPHEMTENCFWIKVNENKYENVDLLCKLENTFCCQQKERREEEDIE
EKKSIKKKIKELKFLDSKIAQNLSIFLSSFRVPYEEIRMMILEVDETRLAESMIQNLIKHLPDQEQLNSL
SQFKSEYSNLCEPEQFVVVMSNVKRLRPRLSAILFKLQFEEQVNNIKPDIMAVSTACEEIKKSKSFSKLL
ELVLLMGNYMNAGSRNAQTFGFNLSSLCKLKDTKSADQKTTLLHFLVEICEEKYPDILNFVDDLEPLDKA

SKVSVETLEKNLRQMGRQLQQLEKELETFPPPEDLHDKFVTKMSRFVISAKEQYETLSKLHENMEKLYQS
IIGYYAIDVKKVSVEDFLTDLNNFRTTFMQAIKENIKKREAEEKEKRVRIAKELAERERLERQQKKKRLL
EMKTEGDETGVMDNLLEALQSGAAFRDRRKRTPMPKDVRQSLSPMSQRPVLKVCNHENQKVQLTEGSRSH
YNINCNSTRTPVAKELNYNLDTHTSTGRIKAAEKKEACNVESNRKKETELLGSFSKNESVPEVEALLARL
RAL

>DEX0356_022.aa.1
MSLADELLADLEEAAEEEEGGSYGEEEEEPAIEDVQEETQLDLSGDSVKTIAKLWDSKMFAEIMMKIEEY
ISKQAKASEVMGPVEAAPEYRVIVDANNLTVEIENELNIIHKFIRDKYSKRFPELESLVPNALDYIRTVK
ELGNSLDKCKNNENLQQILTNATIMVVSVTASTTQGQQLSEEELERLEEACDMALELNASKHRIYEYVES
RMSFIAPNLSIIIGASTAAKIMGVAGGLTNLSKMPACNIMLLGAQRKTLSGFSSTSVLPHTGYIYHSDIV
QSLPPDLRRKAARLVAAKCTLAARVDSFHESTEGKVGYELKDEIERKFDKWQEPPPVKQVKPLPAPLDGQ
RKKRGGRSVRGPGGPVGMGVMEGRSRRPPPSRLPGAAHPPVPVPQDEGAAGADGDPEAGQPYELRRGQTP
RAPSSTPQPASRHRPLPPATAPPLVLWPWLMSRALPQPPPPRPLFSFPSIQPQSDPRGPWSLCLRCLEPP
RLPIAPGSLAGSSLPRGSLVPCCTAAPSLGPASLLCYPSVIPLVLQDRTQRPPHPIKPVLVPDIPRPTRI
EEDAYQEDLGFSLGHLGKSGSGRVRQTQVNEATKARISKTLQRTLQKQSVVYGGKSTIRDRSSGTASSVA
FTPLQGPGDCEPTGGREEGG

>DEX0356_023.aa.1
RRPYAGTRLPVGSPGLFLNAAAATPRTPSVTGATETVTPSEAPVLAAEPEADKGTVLALTENNFDDTIAE
GITFIKFYAPWCGHCKTLAPTWEELSKKEFPGLAGVKIAEVDCTAERNICSKYSVRGYPTLLLFRGGKKV
SEHSGGRDLDSLHRFVLSQAKDEL

>DEX0356_023.aa.2
TDQHSIESTVTLRSEERQYCHKATFPELVGR

>DEX0356_023.aa.3
MVTLEFPYERRDVKTKDRCQWAALALVCTAVAAVDASVL

>DEX0356_024.aa.1
MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKLKDPANFQYPA
ESVLAYKEGHLSPDIVAEQKKLEAADLVIFQVWGDIGRGVRDICVLIVLDMYLISYEILNSYPLYCILCK

>DEX0356_025.aa.1
FVSLCSGSSSCRSLLFFFRFVLIRWSFPLLSSSFSSSLFVVLFRRCGLVRFSRSVLASVLLALLLLSSCV
RFPVACLSFSLLLVICFSLFLLFLSPVSPSFLVSSSPFLLFACACLARSVFFCLCFCRVRLSLVFFGLLF
LFSPLRSLLFSVLRASVPFVFFVFFASFRSLRSSSSVPLLSSFLPLSPFLLLWLPSLAVLPLRLPLLPSV
VSRCCSCVLLCVLVLFWFLVGGCVVCALCVLFVVFVRSWCTAEKSHHQRTSFNRLIVGASPEGLRAGRSG
GCSRLLFFAPWALSKRSRYLALEGTLAPPFFFCMSTFAFIE

>DEX0356_026.aa.1
FESVPCKGEYETCNSDNLNHAVLAVGYGIQKGNKHWIIKNSWGENWGNKGYILMARNKNNACGIANLASF
PKM

>DEX0356_027.aa.1
MEENLISMREDHSFHVRYRMEASCLELALEGERLCKSGDCRAGVSFFEAAVQVGTEDLKTLSAIYSQLGN
AYFYLHDYAKALEYHHHDLTLARTIGDQLGEAKASGNLGNTLKVLGNFDEAIVCCQRHLDISRELNDKVG
EARALYNLGNVYHAKGKSFGCPGPQDVGEFPEEVRDALQAAVDFYEENLSLVTALGDRAAQGRAFGNLGN
THYLLGNFRDAVIAHEQRLLIAKEFGDKAAERRAYSNLGNAYIFLGEFETASEYYKKTLLLARQLKDRAV
EAQSCYSLGNTYTLLQDYEKAIDYHLKHLAIAQELNDRIGEGRACWSLGNAYTALGNHDQAMHFAEKHLE
ISREVGDKSGELTARLNLSDLQMVLGLSYSTNNSIMSENTEIDSSLNGVRPKLGRRHSMENMELMKLTPE
KVQNWNSEILAKQKPLIAKPSAKLLFVNRLKGKKYKTNSSTKVLQDASNSIDHRIPNSQRKISADTIGDE
GFFDLLSRFQSNRMDDQRCCLQEKNCHTASTTTSSTPPKMMLKTSSVPVVSPNTDEFLDLLASSQSRRLD
DQRASFSNLPGLRLTQNSQSVLSHLMTNDNKEADEDFFDILVKCQGSRLDDQRCAPPPATTKGPTVPDED
FFSLILRSQGKRMDEQRVLLQRDQNRDTDFGLKDFLQNNALLEFKNSGKKSADH

>DEX0356_028.aa.1
MARSLSIMFVPNSWLNGYHKVAMNGKARKEAVQTAAKELLKFVNRSPSPFHAVAECRNRLLQAGFSELKE
TEKWNIKPESKYFMTRNSSTIIAFAVGGQYVPGNGFSLIGAHTDSPCLRVKRRSRRSQVGFQQVGVETYG

GGIWSTWFDRDLTLAGRVIVKCPTSGRLEQQLVHVERPILRIPHLAIHLQRNINENFGPNTEMHLVPILA
TAIQEELEKGTPEPGPLNAVDERHHSVLMSLLCAHLGLSPKDIVEMELCLADTQPAVLGGAYDEFIFAPR
LDNLHSCFCALQALIDSCAGPGSLATEPHVRMVTLYDNEEVGSESAQGAQSLLTELVLRRISASCQHPTA
FEEAIPKSFMISADMAHAVHPNYLDKHEENHRPLFHKGPVIKVNSKQRYASNAVSEALIREVANKVKVPL
QDLMVRNDTPCGTTIGPILASRLGLRVLDLGSPQLAMHSIREMACTTGVLQTLTLFKGFFELFPSLSHNL
LVD

>DEX0356_028.aa.2
MSGHSPTRGAMQVAMNGKARKEAVQTAAKELLKFVNRSPSPFHAVAECRNRLLQAGFSELKETEKWNIKP
ESKYFMTRNSSTIIAFAVGGQYVPGNGFSLIGAHTDSPCLRVKRRSRRSQVGFQQVGVETYGGGIWSTWF
DRDLTLAGRVIVKCPTSGRLEQQLVHVERPILRIPHLAIHLQRNINENFGPNTEMHLVPILATAIQEELE
KGTPEPGPLNAVDERHHSVLMSLLCAHLGLSPKDIVEMELCLADTQPAVLGGAYDEFIFAPRLDNLHSCF
CALQALIDSCAGPGSLATEPHVRMVTLYDNEEVGSESAQGAQSLLTELVLRRISASCQHPTAFEEAIPKS
FMISADMAHAVHPNYLDKHEENHRPLFHKGPVIKVAKAPTCRPYGLLQPFGDAIKTLHQRALKPANLPST
FKSTADLAFNIVLLLKPPSIAEPRVNYTKVYLLRPLRAGTHPCQGGTAAVAGDAPHSSPILSATRAAQKH
KQQGARRRM

>DEX0356_029.aa.1
MHLITDDEAPYRTPPPSNAHVQRSINALIDY

>DEX0356_030.aa.1
TSTAQCAGIRLSSGARAGTVNNDEGEWSGPPPECRGKSLTSKVPPTVQKPTTVNVPTTEVSPTSQKTTTK
TTTPNAQATRSTSAATTLS

>DEX0356_031.aa.1
MLKTASAALTWGLGPDPAWSHPAQKTGPPVPVTHCSYGARGFWCWGPPCRWGRRRSRSCREDQKPVMDDQ
RDLISNNEQLPMLGRRPGAPESKCSRGALYTGFSILVTLLLAGQATTAYFLYQQQGRLDKLTVTSQNLQL
ENLRMKLPKPPKPVSKMRMATPLLMQALPMGALPQGPMQNATKYGNMTEDHVMHLLQNADPLKVYPPLKG
SFPENLRHLKNTMETIDWKVFESWMHHWLLFEMSRHSLEQKPTDAPPKVLTKCQEEVSHIPAVHPGSFRP
KCDENGNYLPLQCYGSIGYCWCVFPNGTEVPNTRSRGHHNCSGKQWHCASVRGPGRTRKVEGQEVPSEAH
GTRTPRMAAPGGSDVVMRPALYPSTHLCIPICPFIISPLTCTHICSLIRQSFHL

>DEX0356_031.aa.2
ILHGPLGQGSHGQRLHQNADPLKVYPPLKGSFPENLRHLKNTMETIDWKVFESWMHHWLLFEMSRHSLEQ
KPTDAPPKESLELEDPSSGLGVTKQDLGPVPM

>DEX0356_032.aa.1
MSAHLQWMVVRNCSSFLIKRNKQTYSTEPNNLKARNSFRYNGLIHRKTVGVEPAADGKGVVVVIKRRSGQ
RKPATSYVRTTINKNARATLSSIRHMIRKNKYRPDLRMAAIRRASAILRSQKPVMVKRKRTRPTKSS

>DEX0356_033.aa.1
MEGVLCPAGWWLRPQACAQGRWQHSEATSTAELAVEIVYVVPSVAESLPAPHWLSTQGFQNLEKGAVFWS
WSCGVPCASMGWVCAQAALPWKCGCGYGLGVWDPRQPKWEQGRPVGKGGSGLVGSAAPRCPFSVQRGSDE
LFSTCVTNGPFIMSSNSASAANGNDSKKFKGDSRSAGVPSRVIHIRKLPIDVTEGEVISLGLPFGKVTNL
LMLKGKNQAFIEMNTEEAANTMVNYYTSVTPVLRGQPIYIQFLCALSTAPASCPGGSPRIPFP

>DEX0356_033.aa.2
MKAWFFPFSIRRLVTFPKGSPREMTSPSVTSMGLFRRLHSVPRGSALCAMDGIVPDIAVGTKRGSDELFS
TCVTNGPFIMSSNSASAANGNDSKKFKGDSRSAGVPSRVIHIRKLPIDVTEGEVISLGLPFGKVTNLLML
KGKNQVPEPRFSGVLITVQAGTRRAQRSLPPQQAFIEMNTEEAANTMVNYYTSVTPVLRGQPIYIQFSNH
KELKTDSSPNQARAQAALQAVNSVQSGNLALAASAAAVDAGMAMAGQSPVLRIIVENLFYPVTLDVLHQI
FSKFGTVLKIITFTKNNQFQALLQYADPVSAQHAKLSLDGQNIYNACCTLRIDFSKLTSLNVKYNNDKSR
DYTRPDLPSGDSQPSLDQTMAAAAFASPYAGAGFPPTFAIPQAAGYSNAWSWFPSDCMPTPPSQAAPHPRR
QPGRTGHSSARSGALPGRGDATSTEQAWPGQW

>DEX0356_033.aa.3
MKAWFFPFSIRRLVTFPKGSPREMTSPSVTSMGLFRRLHSVPRGSALCAMDGIVPDIAVGTKRGSDELFS
TCVTNGPFIMSSNSASAANGNDSKKFKGDSRSAGVPSRVIHIRKLPIDVTEGEVISLGLPFGKVTNLLML
KGKNQVPEPRFSGVLTPCRRGRGGPSAHCLPNEAFIEMNTEEAANTMVNYYTSVTPVLRGQPIYIQFSNH
KELKTDSSPNQARAQAALQAVNSVQSGNLALAASAAAVDAGMAMAGQSPVLRIIVENLFYPVTLDVLHQI

FSKFGTVLKIITFTKNNQFQALLQYADPVSAQHAKLSLDGQNIYNACCTLRIDFSKLTSLNVKYNNDKSR
DYTRPDLPSGDSQPSLDQTMAAAFGYRGCPTRRQCAEWSIRAAQSGAPAARARPCTRVMHLLLSARPAQH
GPVSPTSGPRPPLEQRICPRRLCRGRPPAGLGPIPQHSARSLVEVVGAMISVSFISRCTWYNLSLSVCRS
WFPSHLCHSSSCRVFKRLVLVPQRLHAHTTFPGSSASTAAAWADWALECQVRGPSRERGRHVHRAGLARA
VVGQGCGGPRVDGASQKPLIAQSLWSGSLQAPPGEREDRQWLEGCL

>DEX0356_033.aa.4
MKAWFFPFSIRRLVTFPKGSPREMTSPSVTSMGLFRRLHSVPRGSALCAMDGIVPDIAVGTKRGSDELFS
TCVTNGPFIMSSNSASAANGNDSKKFKGDSRSAGVPSRVIHIRKLPIDVTEGEVISLGLPFGKVTNLLML
KGKNQVPEPRFSGVLITVQAGTRRAQRSLPPQQAFIEMNTEEAANTMVNYYTSVTPVLRGQPIYIQFSNH
KELKTDSSPNQARAQAALQAVNSVQSGNLALAASAAAVDAGMAMAGQSPVLRIIVENLFYPVTLDVLHQI
FSKFGTVLKIITFTKNNQFQALLQYADPVSAQHAKLSLDGQNIYNACCTLRIDFSKLTSLNVKYNNDKSR
DYTRPDLPSGDSQPSLDQTMAAAFGLSVPNVHGALAPLAIPSAAAAAAAAGRIAIPGLAGAGNSVLLVSN
LNPEASTVTCSA

>DEX0356_034.aa.1
MRSVQAGLSSQESLSPVLSLSPDSMSFTTRSTFSTNYRSLGSVQAPSYGARPVSSAASVYAGAGGSGSRI
SVSRSTSFRGGMGSGGLATGIAGGLAGMGGIQNEKETMQSLNDRLASYLDRVRSLETENRRLESKIREHL
EKKGPQVRDWSHYFKIIEDLRAQIFANTVDNARIVLQIDNARLAADDFRVKYETELAMRQSVENDIHGLR
KVIDDTNITRLQLETEIEALKEELLFMKKNHEEDRTAATSPDASSGLPRRQNAPNTGTGRPGDSARDEWS
RPRGEGRQNGEPRPKRSVNEPVERPGSDDSGHGGT

>DEX0356_034.aa.2
MSFTTRSTFSTNYRSLGSVQAPSYGARPVSSAASVYAGAGGSGSRISVSRSTSFRGGMGSGGLATGIAGG
LAGMGGIQNEKETMQSLNDRLASYLDRVRSLETENRRLESKIREHLEKKGPQVRDWSHYFKIIEDLRAQI
FANTVDNARIVLQIDNARLAADDFRVKYETELAMRQSVENDIHGLRKVIDDTNITRLQLETEIEALKEEL
LFMKKNHEEEVKGLQAQIASSGVDRGGRCPQISGPRQDHGRHSGPNMTSWLGRTERS

>DEX0356_034.aa.3
MRNLKASLENSLREVEARYALQMEQLNGILLHLESELAQTRAEGQRQAQEYEALLNIKVKLEAEIATYRR
LLEDGEDFNLGDALDSSNSMQTIQKTTTRRIVDGKVVSETNDTKVLRH

>DEX0356_035.aa.1
MLETKWSLLQQQKTARSNMDNMFESYINNLRRQLETLGQEKLKLEAELGNMQGLVEDFKNKSEQLPPSTQ
LKSPCSPPPLGTLGLSPHGPLLWAVQSYLIRVTT

>DEX0356_036.aa.1
MGGAERGRAPAFLLRSAPVSAGGGGAYITCAAPLLREDVACSLAPGESPRLPRELGERNWRRSTPGGSAL
QHEPISSRSRSRRPGRGAASEDENGDDQGFQEGEDELGDEEEGAGDENGHGEQQPQPPATQQQQPQQQRG
AAKEAAGKSSGPTSLFAVTVAPPGARQGQQQAGGDGKTEQKGGDKKRGVKRPREDHGRGYFEYIEENKYS
RAKSPQPPVEEEDEHFDDTVVCLDTYNCDLHFKISRDRLSASSLTMESFAFLWAGGRASYGVSKGKVCFE
MKVTEKIPVRHLYTKDIDIHEVRIGWSLTTSGMLLGEEEFSYGYSLKGIKTCNCETEDYGEKFDENDVIT
CFANFESDEVELSYAKNGQDLGVAFKISKEVLAGRPLFPHVLCHNCAVEFNFGQKEKPYFPIPEEYTFIQ
NVPLEDRVRGPKGPEEKKDCEVVMMIGLPGAGKTTWVTKHAAENPGKYNILGTNTIMDKMMVAGFKKQMA
DTGKLNTLLQRAPQCLGKFIEIAARKKRNFILDQTNVSAAAQRRKMCLFAGFQRKAVVVCPKDEDYKQRT
QKKAEVEGKDLPEHAVLKMKGNFTLPEVAECFDEITYVELQKEEAQKLLEQYKEESKKALPPEKKQNTGS
KKSNKNKSGKNQFNRGGGHRGRGGFNMRGGNFRGGAPGNRGGYNRRGNMPQRGGGGGGSGGIGYPYPRAP
VFPGRGSYSNRGNYNRGGMPNRGNYNQNFRGRGNNRGYKNQSQGYNQWQQGQFWGQKPWSQHYHQGYY

>DEX0356_037.aa.1
MIFRDFIFAQVISKYEDIFFVFFQVGCLKFFPFFIPLPKANEKKVDQPPEAKKPKIKVVNVELPIEANLV
WQLGKDLLNMYIETEGKMIMQDKLEKERNDAKNAVEEYVYEFRDKLCGPYEKFICEQDHQNFLRLLTETE
DWLYEEGEDQAKQAYVDKLEELMKIGTPVKVRFQEAEERPKMFEELGQRLQHYAKIAADFRNKDEKYNHI
DESEMKKVEKSVNEVMEWMNNVMNAQAKKSLDQDPVVRAQEIKTKIKELNNTCEPVVTQPKPKIESPKLE
RTPNGPNIDKKEEDLEDKNNFGAEPPHQNGECYPNEKNSVNMDLD

>DEX0356_038.aa.1
MKVQNRRNAHLSSFDRFAFLWEMQFRVTCCRRYVVDHLDIIAEDELLDQLLSHTASLSDLLQNSVVATAQ
TRLRSYSCASLRFSSATMSDKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKQAGES

>DEX0356_039.aa.1
MCLGVCSFRCSDMSRVSSFQWVRGLTDFRNEAAYAHAPQKSSPSPHSTQEVLLSSPVTSYWGG

>DEX0356_040.aa.1
MNGEYYQESNGPTDSYAAISQVDRLQSEPESIRKWREEQMERLEALDANSRKQEAEWKEKAIKELEEWYA
RQDEQLQKTKANNRAAEEAFVNDIDESSPGTEWERVARLCDFNPKSSKQAKDVFPPCAQSSSPSSRPLVL
EAPVDLLVLPRGLRWVLGWPLLA

>DEX0356_040.aa.2
MAELDPFGAPAGAPGGPALGNGVAGAGEEDPAAAFLAQQESEIAGIENDEAFAILDGAAPPGPSRTASRR
GVRMQESNGPTDSYAAISQVDRLQSEPESIRKWREEQMERLEALDANSRKQEAEWKEKAIKELEEWYARQ
DEQLQKTKANNRAAEEAFVNDIDESSPGTEWERVARLCDFNPKSSKQAKDVFPPCAQSSSPSSRPRWCTE
EPPCGNTTSAIS

>DEX0356_040.aa.3
SSRSRAAARDLEPDAWGACPPLPVGTTAVAAGRGVGGSVGFCLTVGVRAVQLPAMAELDPFGAPAGAPGG
PALGNGVAGAGEEDPAAAFLAQQESEIAGIENDEAFAILDGAAPPGPSRTASRRGVRMPILGSKKQSGKK
RQ

>DEX0356_040.aa.4
MNGEYYQESNGPTDSYAAISQVDRLQSEPESIRKWREEQMERLEALDANSRKQEAEWKEKAIKELEEWYA
RQDEQLQKTKANNRVADEAFYKQPFADVIGYVQGSRRSLCK

>DEX0356_040.aa.5
MNGEYYQESNGPTDSYAAISQVDRLQSEPESIRKWREEQMERLEALDANSRKQEAEWKEKAIKELEEWYA
RQDEQLQKTKANNRAAEEAFVNDIDESSPGTEWERVARLCDFNPKSSKQAKDVFPPCAQSSSPSSRPVGD
EEHLWKTNLIFIYQGSLSMIYVEVLENLCSSVLLVGFSKGSGKRGRKGAKRWFKFGRKKKGLGCFVGLKF
I

>DEX0356_040.aa.6
MNGEYYQESNGPTDSYAAISQVDRLQSEPESIRKWREEQMERLEALDANSRKQEAEWKEKAIKELEEWYA
RQDEQLQKTKANNRAAEEAFVNDIDESSPGTEWERVARLCDFNPKSSKRARCFPMAQTHSSRRPVCLRAP
VEPTLIFIYQ

>DEX0356_041.aa.1
MKALIVLGLVLLSVTVQGKVFERCELARTLKRLGMDGYRGISLANWMCLAKWESGYNTRATNYNAGDRST
DYGIFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIADAVACAKRVVRDPQGIRAWVAWRNRCLSFFSS
FCLSHIKGVGIK

>DEX0356_042.aa.1
SEGGAWTPGSEPTGAQACEFEAPSHHPPPQVVYIACSFTTVWLIYSKFKATYDGNHDTFRVEFLVVPTAI
LAFLVNHDFTPLEILWTFSIYLESVAILPQLFMVSKTGEAETITSHYLFALGVYRTLYLFNWIWRYHFEG
FFDLIAIVAGLVQTVLYCDFFYLYITKVLKGKKLSLPA

>DEX0356_043.aa.1
MLVAITVLIYPNSIQEYIRQLPPNFPYRDDVMSVNPTCLVLIILLFISIILTFKGYLISCVWNCYRYING
RNSSDVLVYVTSNDTTVLLPPYDDATVNGAAKEPPPPYVSA

>DEX0356_043.aa.2
MKMVAPWTRFYSNSCCLCCHVRTGTILLGVWYLIINAVVLLILLSALADPDQYNFSSSELGGDFEFMDDA
NMCIAIAISLLMILICAMATYGAYKQRAAWIIPFFCYQIFDFALNMLVAITVLIYPNSIQEYIRQLPPNF
PYRDDVMSVNPTCLVLIILLFISIILTFKGYLISCVWNCYRYINGRNSSDVLVYVTSNDTTGCTR

>DEX0356_044.aa.1
MADAGIRRVVPSDLYPLVLGFLRDNQLSEVANKFAKATGATQQDANASSLLDIYSFWLNRSAKVPERKLQ
ANGPVAKKAKKKASSSDSEDSSEEEEEVQGPPAKKAAVPAKRVGLPPGKAAAKASESSSSEESSDDDDEE
DQKKQPVQKGVKPQAKAAKAPPKKAKSSDSDSDSSSEDEPPKNQKPKITPVTVKAQTKAPPKPARAAPKI
ANGKAASSSSSSSSSSSSDDSEEEKAAATPKKVWTITSVRAEVTRAVMCVCLPSLAGLVGSAFSCGESIF
FLMQTVPKKQVVAKAPVKAATTPTRKSSSSEDSSSDEEEEQKKPMKNKPGPYSSVPPPSAPPPKKSLGTQ
PPKKAVEKQQPVESSEDSSDESDSSSEEEKKPPTKAVVSKATTKPPPAKKAAESSSDSSDSDSSEDDEAP

SKPAGTTKNSSNKPAVTTKSPAVKPAAAPKQPVGGGQKLLTRKADSSSSEEESSSSEEEKTKKMVATTKP
KATAKAALSLPAKQAPQGSRDSSSDSDSSSSEEEEEKTSKSAVKKKPQKVAGGAAPSKPASAKKGKAESS
NSSSSDDSSEEEEEKLKGKGSPRPQAPKANGTSALTAQNGKAAKNSEEEEEEKKKAAVVVSKSGSLKKRK
QNEAAKEAETPQAKKIKLQTPNTFPKRKKGEKRASSPFRRVREEEIEVDSRVADNSFDAKRGAAGDWGER
ANQVLKFTKGKSFRHEKTKKKRGSYRGGSISVQVNSIKFDSE

>DEX0356_044.aa.2
LSGHLLLKQSGPTQPERTAFLSNPQPLIFANLKHYQKQQSLPSLCSTSDPDTPLGAPSSSDSDSSSSEEE
EEKTSKSAVKKKPQKVAGGAAPSKPASAKKGKAESSNSSSSDDSSEEEEEKLKGKGSPRPQAPKANGTSA
LTAQNGKAAKNSEEEEEEKKKAAVVVSKSGSLKKRKQNEAAKEAETPQAKKIKLQTPNTFPKRKKGEKRA
SSPFRRVREEEIEVDSRVADNSFDAKRGAAGDWGERANQVLKFTKGKSFRHEKTKKKRGSYRGGSISVQV
NSIKFDSE

>DEX0356_045.aa.1
MDSVVAARYKDSAQRDDMIFEDCGDVPSEPKERGDFNHERGERKVCRFKLEWLGNCSGLNDETYGYKEGK
PCIIIKLNRVLGFKPKPPKNESLETYPVMKYNPNVLPVQCTGKRDEDKDKVGNVEYFGLGNSPGFPLQYY
PYYGKLLQPKYLQPLLAVQFTNLTMDTEIRIECKAYGENIGYSEKDRFQGRFDVKIEVKS

>DEX0356_046.aa.1
MILVQILKQEWPKHWPTFISDIVGASRTSESLCQNNMVILKLLSEEVFDFSSGQITQVKSKHLKDSMCNE
FSQIFQLCQFVMENSQNAPLVHATLETLLRFLNWIPLGYIFETKLISTLIYKFLNVPMFRNVSLKCLTEI
AGVSVSQYEEQFVTLFTLTMMQLKQMLPLNTNIRLAYSNGKDDEQNFIQNLSLFLCTFLKEHDQLIEKRL
NLRETLMEALHYMLLVSEVEETEIFKICLEYWNHLAAELYRESPFSTSASPLLSGSQHFDVPPRRQLYLP
MLFKVTERLVECSSCCILWFLRSESKYFYICVNKLAIKREPNNFSMSVENQNMKGVESRTLILKSVVLLS
VSTLVVISLGKFIATCQSTKPESRNETQETPVTEVGEKNHIKTHLNNYKAIFVELQWKKNFFFWRQGLAL
WLRLECSGVVIAHYNLELLDYKQFSCVSLPSNWLQAHTTMPEQILNFLVLLHLYLHSFFCCALSPSIFIV
LEIGCHFVLLKLLSNSWPQQFFLPRPSGIVGIIGMNPYTWLACGFFGFVCFELSFYTNDFSDLTIFFLID
ILLSMVRLLMVSRMAKPEEVLVVENDQGEVVREFMKDTDSINLYKNMRETLGKLINTVNLYLYKSILGWK
YILGNVLVVSKYFLNVLLLINKILLIIAFETMFFFVVYLTHLDYVDTERIMTEKLHNQVNGTEWSWKNLN
TLCWAIGSISGAMHEEDEKRFLVTVIKDLLGLCEQKRGKDNKAIIASNIMYIVGQYPRFLRAHWKFLKTV
VNKLFEFMHETHDGVQDMACDTFIKIAQKCRRHFVQVQVGEVMPFIDEILNNINTIICDLQPQQVHTFYE
AVGYMIGAQTDQTVQEHLIEKYMLLPNQVWDSIIQQATKNVDILKDPETVKQLGSILKTNVRACKAVGHP
FVIQLGRIYLDMLNVYKCLSENISAAIQANGEMVTKQPLIRSMRTVKRETLKLISGWVSRSNDPQMVAEN
FVPPLLDAVLIDYQRNVPAARREPEVLSTMAIIVNKLGGHITAEIPQIFDAVFECTLNMINKDFEEYPEHR
TNFFLLLQAVNSHCFPAFLAIPPTQFKLVLDSIIWAFKHTMRNVADTGLQILFTLLQNVAQEEAAAQSFY
QTYFCDILQHIFSVVTDTSHTAGLTMHASILAYMFNLVEEGKISTSLNPGNPVNNQIFLQEYVANLLKSA
FPHLQDAQVKLFVTGLFSLNQDIPAFKEHLRDFLVQIKEFAGEDTSDLFLEEREIALRQADEEKHKRQMS
VPGIFNPHEIPEEMCD

>DEX0356_047.aa.1
MELHYLTHFQIPDRISTVRRFTLLLSFLLRRLVTLISQQATLLASNEAFKKQAESASEAAKKYMEENDQL
KKGAAVDGGKLDVGNAEVKLEEENRSLKADLQKLKDELASTKQKLEKAENQVLAMRKQSEGLTKEYDRLL
EEHAKLQAAVDGPMDKKEE

>DEX0356_047.aa.2
MGAEASSSWCPGTALPEERLSVKRASEISGFLGQGSSGEAALDVLTHVLEGAGNKLTSSCGKPSSNRMSL
QWTAVATFLYAEVFVVLLLCIPFISPKRWQKIFKSRLVELLVSYGNTFFVVLIVILVLLVIDAVREIRKY
DDVTEKVNLQNNPGAMEHFHMKLFRAQRNLYIAGFSLLLSFLLRRLVTLISQQATLLASNEAFKKQAESA
SEAAKKYMEENDQLKKGAAVDGGKLDVGNAEVKLEEENRSLKADLQKLKDELASTKQKLEKAENQVLAMR
KQSEGLTKEYDRLLEEHAKLQAAVDGPMDKKEE

>DEX0356_047.aa.3
MSLQWTAVATFLYAEVFVVLLLCIPFISPKRWQKIFKSRLVELLVSYGNTFFVVLIVILVLLVIDAVREI
RKYDDVTEKVNLQNNPGAMEHFHMKLFRAQRNLYIAGFSLLLSFLLRRLVTLISQQATLLASNEAFKKQA
ESASEAAKKYMEENDQLKKGAAVDGGKLDVGNAEVKLEEENRSLKADLQKLKDELASTKQKLEKAENQVL
AMRKQSEGLTKEYDRLLEEHAKLQSPASPSSTSVCALLLPPFPSTAHSSSSRPLSTLSKHITGDLIATRS
ECVCCHPAWPGQAWHSLGFHA

>DEX0356_047.aa.4
SRPSAGPLPHHRDVGRPRGLEKRPSPARQSPAPPRPRARSTPPCSPGARGASSSPARSAGPPLRGGVGSG

GGRSLRGLARAGGRCPAGASGFRPRYEGRGRGCCGRVRLLRRGLHVDCGKLGNKLTSSCGKPSSNRMSLQ
WTAVATFLYAEVFVVLLLCIPFISPKRWQKIFKSRLVELLVSYGNTFFVVLIVILVLLVIDAVREIRKYD
DVTEKVNLQNNPGAMEHFHMKLFRAQRNLYIAGFSLLLSFLLRRLVTLISQQATLLASNEAFKKQAESAS
EAAKKYMEENDQLKKGAAVDGGKLDVGNAEVKLEEENRSLKADLQKLKDELASTKQKLEKAENQVLAMRK
QSEGLTKEYDRLLEEHAKLQGQPSRIAAPRAGEQSRMRQPLRV

>DEX0356_047.aa.5
SRPSAGPLPHHRDVGRPRGLEKRPSPARQSPAPPRPRARSTPPCSPGARGASSSPARSAGPPLRGGVGSG
GGRSLRGLARAGGRCPAGASGFRPRYEGRGRGCCGRVRLLRRGLHVDCGKLGNKLTSSCGKPSSNRMSLQ
WTAVATFLYAEVFVVLLLCIPFISPKRWQKIFKSRLVELLVSYGNTFFVVLIVILVLLVIDAVREIRKYD
DVTEKVNLQNNPGAMEHFHMKLFRAQRNLYIAGFSLLLSFLLRRLVTLISQQATLLASNEAFKKQAESAS
EAAKTRPRPR

>DEX0356_048.aa.1
IXSPFSYTPPFPPTGGAALSARGIXGPEXEPAAAAVARGKREEAATGEKKRSRHRREVDTQAAAAARXAR
PPXRPSPRPVHKAARLRKGVRTKGQSLQ

>DEX0356_049.aa.1
MDMSLELDIIKLNRSQRGGWGGCRGLRAVCRLPVAGRVWWVLQAAAQVNRGGGPIRNRPGIALRAAVEGG
RNRPAPYSRPKQLPDKWQHDLFDSGFGGGAGVETGGKLCVVSNLDIGVSDADIQELFAEFGTLKKAAVHY
DRSGRSLGTADVHFERKADALKAMKQYNGVPLDGRPMNIQLVTSQIDAQRRPAQSVNRGGMTRNRGAGGM
GELVEAPRRRHCSDGFCGRGRGAGRNSKQQLSAEELDAQLDAYNARMDTS

>DEX0356_050.aa.1
MTVAEIKGTTGRTYVGSHAQGRSSKCLKKTKTENPTDPHQTRWTRSGQRLHRGTQSVRQHAWSMHGPKSK
NAKHSMDQHTHGQCHVDIVTRCMFICTGQTVTGHHPTSTAAETSMEMINSVRATVAQRESMCHLREAAYT
VVPTGIAHHVLALDESKAKLSSDVLTLLIKQYCRESGVRNLQKQVEKVLRKSAYKIVSGEAESVEVTPEN
LQDFVGKPVFTVERMYDVTPPGVVMGLAWTAMGGSTLFVETSLRRPQDKDAKGDKDGSLEVTGQLGEVMK
ESARIAYTFARAFLMQHAPANDYLVTSHIHLHVPEGATPKDGPSAGCTIVTALLSLAMGRPVRQNLAMTG
EVSLTGKILPVGGIKEKTIAAKRAGVTCIVLPAENKKDFYDLAAFITEGLEVHFVEHYREIFDIAFPDEQ
AEALAVER

>DEX0356_051.aa.1
MSLAILEEVIMWDPRVSRPCLPLGHVATQYFAEQGHASVGWAEYPEEDLKRTMMACGGSNPDQCEGSVSR
LCLARCQVFEETQLGGERYNLFDWLPQAKTCTFILRGGTEQFMEETERSLHDAIMIVMRAIKNDSVVAGG
GAIEMELSKYLRDYSRTIPGKQQLLIGAYAKALEIIPRQLCDNAGFDATNILNKLRARHAQGGTWYGVDI
NNEDIADNFEAFVWEPAMVRINALTAASEAAVPDRVL

>DEX0356_052.aa.1
MAMDLCRQDPECEFYFSXDADAVLTNXQTLXXLIEEXRGVXAPMXXXXGKXWSNXWGXLSXDEYYARSXD
XVELVXRKDVGVWNVPYIXXAYVXGGDTXXMELPQRDVFSGSDXDXDMAFCEXFGXKGIFXHLSYQHEFG
RLLATSRYDTEHLHPDLWQIFDNPVDWKEQYIHENYSRALEGKDIVEQPCPDVYWFPLLSEQMCDELVAE
MEHYGQWSGGRHEDSRLAGGYENVPTVDXPHEAGGVRGPVAAAAADVCGPMTXTCFRFT

>DEX0356_053.aa.1
MKMVAPWTRFYSNSCCLCCHVRTGTILLGVWYLIINAVVLLILLSALADPDQYNFSSSELGGDFEFMDDA
NMCIAIAISLLMILICAMATYGAYKQRAAWIIPFFCYQIFDFALNMLVAITVLIYPNSIQEYIRQLPPNF
PYRDDVMSVNPTCLVLIILLFISIILTFKGYLISCVWNCYRYINGRNSSDVLVYVTSNDTTVLLPPYDDA
TVNGACQGAPAPYVLPKPSSGRT

>DEX0356_054.aa.1
MDSSGARAGTVCTLHGIHRGHQKPGLFLSRSSNVFSSDSHHLQLVPTQCCHSRLTQVLRPSTSSAARAHL
RLDCTI

>DEX0356_054.aa.2
MGFIEATRNQAFSCPEVPSFSSDSHHLQLVPTQCCHSRLTQVLRPSTSSAARAHLRLDCTI

>DEX0356_055.aa.1
ADETSEKEQQEAIEHIDEVQNEIDRLNEQASEEILKVEQKYNKLRQPFVQKRSELIAKIPTFLGNNICQP
STSVCTALREEDEEALHYLTRVEVTEFEDIKSGYRIDFDFDGNPYFENKVLSKEFHLNESGDPSSKSTEI

KWKSGKDLTKRSSQTQNKASRKRQHEEPESFFTWFTDHSDAGADELGEVIKDDIWPNPLQYYLVPDMDDE
EGEGEEDDDDDEEEEGLEDIAKNGDEDEGDGMMKIDDEVGGTEQGGLKKKFLXXPPX

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising:

(a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 101-194;

(b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-100;

(c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b); or

(d) a nucleic acid molecule having at least 95% sequence identity to the nucleic acid molecule of (a) or (b).

2. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a cDNA.

3. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is genomic DNA.

4. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is an RNA.

5. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a mammalian nucleic acid molecule.

6. The nucleic acid molecule according to claim 5, wherein the nucleic acid molecule is a human nucleic acid molecule.

7. A method for determining the presence of a colon specific nucleic acid (CSNA) in a sample, comprising the steps of:

(a) contacting the sample with the nucleic acid molecule of SEQ ID NO: 1-100 under conditions in which the nucleic acid molecule will selectively hybridize to a colon specific nucleic acid; and

(b) detecting hybridization of the nucleic acid molecule to a CSNA in the sample, wherein the detection of the hybridization indicates the presence of a CSNA in the sample.

8. A vector comprising the nucleic acid molecule of claim 1.

9. A host cell comprising the vector according to claim 8.

10. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of:

(a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and

(b) incubating the host cell under conditions in which the polypeptide is produced.

11. A polypeptide encoded by the nucleic acid molecule according to claim 1.

12. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 101-194 ; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 95% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-100.

13. An antibody or fragment thereof that specifically binds to:

(a) a polypeptide comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 101-194 ; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 95% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-100.

14. A method for determining the presence of a colon specific protein in a sample, comprising the steps of:

(a) contacting the sample with a suitable reagent under conditions in which the reagent will selectively interact with the colon specific protein comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 101-194; and

(b) detecting the interaction of the reagent with a colon specific protein in the sample, wherein the detection of binding indicates the presence of a colon specific protein in the sample.

15. A method for diagnosing or monitoring the presence and metastases of colon cancer in a patient, comprising the steps of:

(a) determining an amount of:

(i) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 101-194;

(ii) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-100;

(iii) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (i) or (ii);

(iv) a nucleic acid molecule having at least 95% sequence identity to the nucleic acid molecule of (i) or (ii);

(v) a polypeptide comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 101-194 ; or

(vi) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 95% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-100 and;

(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the colon specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of colon cancer.

16. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence of:

(a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 101-194;

(b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-100;

(c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b); or

(d) a nucleic acid molecule having at least 95% sequence identity to the nucleic acid molecule of (a) or (b); or

(e) a polypeptide comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 101-194 ; or

(f) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 95% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-100.

17. A method of treating a patient with colon cancer, comprising the step of administering a composition consisting of:

(a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 101-194;

(b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-100;

(c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b);

(d) a nucleic acid molecule having at least 95% sequence identity to the nucleic acid molecule of (a) or (b);

(e) a polypeptide comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 101-194 ; or

(f) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 95% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-100;

to a patient in need thereof, wherein said administration induces an immune response against the colon cancer cell expressing the nucleic acid molecule or polypeptide.

18. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 12.

5

ABSTRACT OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and polypeptides present in normal and neoplastic colon cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions containing the nucleic acid molecules, polypeptides, antibodies, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and non-cancerous disease states in colon, identifying colon tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered colon tissue for treatment and research.